



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I733451 B

(45)公告日：中華民國 110 (2021) 年 07 月 11 日

(21)申請案號：109116174

(22)申請日：中華民國 108 (2019) 年 02 月 18 日

(51)Int. Cl. : **G06N3/06 (2006.01)****G11C16/08 (2006.01)**

(30)優先權：2018/03/14 美國

62/642,867

2018/05/25 美國

15/990,220

2019/01/16 世界智慧財產權組織

PCT/US19/13871

(71)申請人：美商超捷公司 (美國) SILICON STORAGE TECHNOLOGY, INC. (US)

美國

(72)發明人：陳 曉萬 TRAN, HIEU VAN (US) ; 蒂瓦里 維平 TIWARI, VIPIN (US) ; 杜 恩

漢 DO, NHAN (US)

(74)代理人：劉法正；尹重君

(56)參考文獻：

TW 201411509A

US 9286205B2

US 9711197B1

US 2016/0300607A1

審查人員：廖國智

申請專利範圍項數：7 項 圖式數：27 共 56 頁

(54)名稱

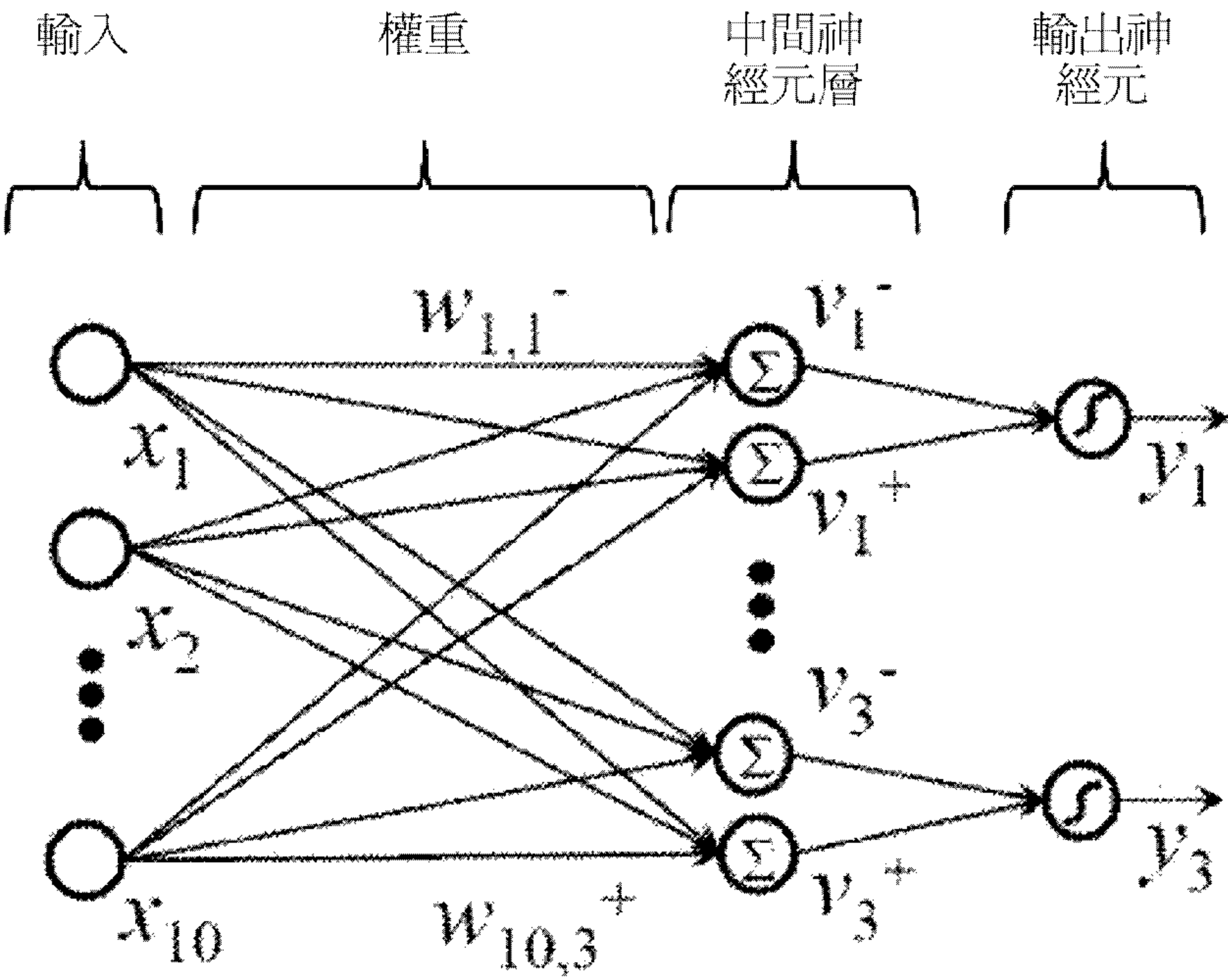
用於偵測類比神經形態記憶體系統中之單元中資料漂移之方法

(57)摘要

揭示一種用於使用人工類神經網路中的一向量矩陣乘法(vector-by-matrix multiplication, VMM)陣列的資料再新方法及設備之多個實施例。揭示一種適用於偵測 VMM 陣列內的快閃記憶體單元中的資料漂移的資料漂移偵測器之各種實施例。

Numerous embodiments of a data refresh method and apparatus for use with a vector-by-matrix multiplication (VMM) array in an artificial neural network are disclosed. Various embodiments of a data drift detector suitable for detecting data drift in flash memory cells within the VMM array are disclosed.

指定代表圖：



(先前技術)

【圖1】



I733451

【發明摘要】

【中文發明名稱】

用於偵測類比神經形態記憶體系統中之單元中資料漂移之方法

【英文發明名稱】

A METHOD OF DETECTING DATA DRIFT IN CELLS IN AN ANALOG NEUROMORPHIC MEMORY SYSTEM

【中文】

揭示一種用於使用人工類神經網路中的一向量矩陣乘法(vector-by-matrix multiplication, VMM)陣列的資料再新方法及設備之多個實施例。揭示一種適用於偵測 VMM 陣列內的快閃記憶體單元中的資料漂移的資料漂移偵測器之各種實施例。

【英文】

Numerous embodiments of a data refresh method and apparatus for use with a vector-by-matrix multiplication (VMM) array in an artificial neural network are disclosed. Various embodiments of a data drift detector suitable for detecting data drift in flash memory cells within the VMM array are disclosed.

【指定代表圖】 圖1

【代表圖之符號簡單說明】

(無)

【特徵化學式】

(無)

【發明說明書】

【中文發明名稱】

用於偵測類比神經形態記憶體系統中之單元中資料漂移之方法

【英文發明名稱】

A METHOD OF DETECTING DATA DRIFT IN CELLS IN AN ANALOG NEUROMORPHIC MEMORY SYSTEM

【技術領域】

【0001】優先權主張

本申請案主張於 2018 年 3 月 14 日申請之美國臨時專利申請案第 62/642,867 號之優先權，其名稱為「Method and Apparatus for Data Refresh for Analog Neuromorphic Memory in Artificial Neural Network」，以及主張於 2018 年 5 月 25 日申請之美國專利申請案第 15/990,220 號之優先權，其名稱為「Method and Apparatus for Data Refresh for Analog Non-volatile Memory in Deep Learning Neural Network」。

【0002】 揭示一種用於使用人工類神經網路中的一向量矩陣乘法 (vector-by-matrix multiplication, VMM) 陣列中的快閃記憶體單元的資料再新方法及設備之多個實施例。

【先前技術】

【0003】 人工類神經網路模仿生物神經網路（動物的中樞神經系統，特別係大腦），其用以評估或近似可取決於大數目的輸入且通常未知的功能。人工類神經網路通常包括在彼此間交換訊息之互連的「神經元」的層。

【0004】 圖 1 繪示一人工類神經網路，其中圓形表示神經元的輸入或

層。連接（稱為突觸）係以箭頭表示，並具有可基於經驗調諧的數字權重。此使類神經網路適應輸入並能夠學習。一般而言，類神經網路包括多個輸入的一層。一般而言有一或多個神經元中間層，及提供類神經網路的輸出之一神經元輸出層。各級的神經元基於自突觸接收之資料個別地或共同地做決策。

【0005】 用於高效能資訊處理的人工類神經網路之發展中的主要挑戰之一者係缺少適當的硬體技術。的確，實際的類神經網路依賴非常大數目的突觸，而在神經元之間實現高連接性，亦即，非常高的計算平行性。原理上，此種複雜性可以數位超級電腦或專用圖形處理單元叢集實現。然而，除了高成本以外，此等方法與生物網路相比時也苦於中等的能量效率，主要因為生物網路實施低精確度類比計算而消耗更少的能量。CMOS 類比電路已用於人工類神經網路，但大多數 CMOS 實作的突觸在給定高數目的神經元及突觸情況下已過於龐大。

【0006】 申請人先前在美國專利申請案第 15/594,439 號（其以引用方式併入本文中）中揭示一種人工（類比）類神經網路，其利用一或多個非揮發性記憶體陣列作為突觸。非揮發性記憶體陣列作為類比神經形態記憶體進行操作。該類神經網路裝置包括一第一複數個突觸及一第一複數個神經元，該第一複數個突觸經組態以接收一第一複數個輸入並由此產生一第一複數個輸出，該第一複數個神經元經組態以接收該第一複數個輸出。該第一複數個突觸包括複數個記憶體單元，其中該等記憶體單元之各者包括形成於一半導體基材中之間隔開的源極區域及汲極區域，其中一通道區域在該源極區域與該汲極區域之間延伸；一浮閘，其設置於該通道區域之一

第一部分上方並與該第一部分絕緣；及一非浮閘，其設置於該通道區域之一第二部分上方並與該第二部分絕緣。該複數個記憶體單元之各者經組態來儲存對應於該浮閘上之數個電子的權重值。該複數個記憶體單元經組態以使該第一複數個輸入乘以所儲存的該等權重值，以產生該第一複數個輸出。

【0007】 該類比神經形態記憶體系統中所用的各非揮發性記憶體單元必須經抹除及程式化以在該浮閘中保持一非常特定且精確的電荷量。例如，各浮閘必須保持 N 個不同值之一者，其中 N 係可由各單元指示之不同權重的數目。 N 的實例包括 16、32、及 64。

【0008】 習知快閃記憶體裝置無需再新其等資料，此係因為各快閃記憶體單元可無限期保持其值（一般係限於兩個可能位準或有時四個可能位準，即， $N = 2$ 或 4）。然而，在 VNM 中所使用的快閃記憶體單元中，因為 N 顯著較大，資料隨著時間而變得不正確（即，毀壞）的可能性增加，其係由於資料漂移隨著時間基於溫度及程序變異的改變而自然地發生。

【0009】 所需要的係一種用於偵測 VNM 中的快閃記憶體單元中的資料漂移，以及將該單元中的資料再新至其原始、所欲狀態的機制。

【發明內容】

【0010】 揭示一種用於使用人工類神經網路中的一向量矩陣乘法 (vector-by-matrix multiplication, VMM) 陣列中的快閃記憶體單元的資料再新方法及其設備之多個實施例。

【圖式簡單說明】

【0011】 圖 1 係繪示一人工類神經網路的圖。

【0012】 圖 2 係習知的 2 閘極非揮發性記憶體單元的截面側視圖。

【0013】 圖 3 係習知的 4 閘極非揮發性記憶體單元的截面側視圖。

【0014】 圖 4 係習知的 3 閘極非揮發性記憶體單元的截面側視圖。

【0015】 圖 5 係另一習知的 2 閘極非揮發性記憶體單元的截面側視圖。

【0016】 圖 6 係繪示利用非揮發性記憶體陣列的例示性人工類神經網路之不同級的圖。

【0017】 圖 7 係繪示一向量乘法器矩陣的方塊圖。

【0018】 圖 8 係繪示一向量乘法器矩陣之各種級的方塊圖。

【0019】 圖 9 描繪一向量乘法器矩陣之實施例。

【0020】 圖 10 描繪圖 9 之向量乘法器矩陣之操作電壓。

【0021】 圖 11 描繪一向量乘法器矩陣之另一實施例。

【0022】 圖 12 描繪圖 11 之向量乘法器矩陣之操作電壓。

【0023】 圖 13 描繪一向量乘法器矩陣之另一實施例。

【0024】 圖 14 描繪圖 13 之向量乘法器矩陣之操作電壓。

【0025】 圖 15 描繪一向量乘法器矩陣之另一實施例。

【0026】 圖 16 描繪圖 15 之向量乘法器矩陣之操作電壓。

【0027】 圖 17 描繪一向量乘法器矩陣之另一實施例。

【0028】 圖 18 描繪用於偵測一向量乘法器矩陣中的資料漂移之系統。

【0029】 圖 19 描繪一資料再新方法。

【0030】 圖 20 描繪另一資料再新方法。

【0031】 圖 21 描繪一向量乘法器矩陣中的快速單元之特性。

【0032】 圖 22 描繪一向量乘法器矩陣中的快速單元之特性。

【0033】 圖 23 描繪用於評估資料漂移之機制。

【0034】 圖 24 描繪資料漂移偵測器之實施例。

【0035】 圖 25 描繪資料漂移偵測器之另一實施例。

【0036】 圖 26 描繪圖 25 之實施例之例示性取樣時期。

【0037】 圖 27 描繪資料漂移偵測器之另一實施例。

【實施方式】

【0038】 本發明的人工類神經網路利用 CMOS 技術與非揮發性記憶體陣列的組合。

非揮發性記憶體單元

【0039】 數位非揮發性記憶體係已知。例如，美國專利第 5,029,130 號（「'130 專利」）揭示分離閘非揮發性記憶體單元之一陣列，且係為所有目的以引用方式併入本文中。此一類記憶體單元顯示於圖 2 中。各記憶體單元 210 包括形成在半導體基材 12 中的源極區域 14 及汲極區域 16，其間具有通道區域 18。浮閘 20 形成在通道區域 18 的第一部分上方且與該第一部分絕緣（且控制該第一部分的導電性），及形成在源極區域 16 的一部分上方。字線端子 22（其一般耦接至字線）具有設置在通道區域 18 之第二部分上方且與該第二部分絕緣（且控制該第二部分的導電性）的第一部分，及向上並於浮閘 20 上方延伸的第二部分。浮閘 20 及字線端子 22 藉由閘極氧化物與基材 12 絕緣。位元線 24 耦接至汲極區域 16。

【0040】 記憶體單元 210 係藉由將高正電壓置於字線端子 22 上而抹除（其中將電子從浮閘移除），其導致浮閘 20 上的電子經由富爾諾罕穿隧而從浮閘 20 穿隧通過中間絕緣體至字線端子 22。

【0041】 記憶體單元 210 係藉由將正電壓置於字線端子 22 上及將正電壓置於源極 16 上而程式化（其中將電子置於浮閘上）。電子流將自源極 16 朝向汲極 14 流動。當電子抵達字線端子 22 與浮閘 20 之間の間隙時，其等將加速且變成經加熱的。由於來自浮閘 20 的吸引靜電力，該等經加熱電子的一些將通過閘極氧化物 26 注入至浮閘 20 上。

【0042】 記憶體單元 210 係藉由將正讀取電壓置於汲極 14 及字線端子 22 上（其導通字線端子下方的通道區域）而讀取。若浮閘 20 帶正電荷（亦即電子經抹除並正耦接至汲極 16），則浮閘 20 下方的通道區域部分亦經接通，且電流將跨通道區域 18 流動，其係感測為經抹除或「1」狀態。若浮閘 20 帶負電荷（亦即以電子程式化），則浮閘 20 下方的通道區域部分係大部分或完全斷開，且電流將不會跨通道區域 18 流動（或將有少許流動），其係感測為經程式化或「0」狀態。

【0043】 表 1 描繪可施加至記憶體單元 210 的端子之用於執行讀取、抹除、及程式化操作的典型電壓範圍：

表 1：圖 2 之快閃記憶體單元 210 的操作

	WL	BL	SL
讀取	2 至 3 V	0.6 至 2 V	0 V
抹除	~11 至 13 V	0 V	0 V
程式化	1 至 2 V	1 至 3 μ A	9 至 10 V

【0044】 其它分離閘記憶體單元組態為已知。例如，圖 3 描繪四閘極記憶體單元 310，其包含源極區域 14、汲極區域 16、在通道區域 18 的第

一部分上方的浮閘 20、在通道區域 18 的第二部分上方的選擇閘 28（一般耦接至字線）、在浮閘 20 上方的控制閘 22、及在源極區域 14 上方的抹除閘 30。此組態描述於美國專利第 6,747,310 號中，其針對所有目的以引用方式併入本文中。此處，除了浮閘 20 以外的所有閘係非浮閘，意指其等經電氣連接或可連接至電壓源。程式化係藉由變熱的電子自通道區域 18 將其本身注入至浮閘 20 上來顯示。抹除係藉由自浮閘 20 至抹除閘 30 之電子穿隧來顯示。

【0045】 表 2 描繪可施加至記憶體單元 310 的端子之用於執行讀取、抹除、及程式化操作的典型電壓範圍：

表 2：圖 3 之快閃記憶體單元 310 的操作

	WL/SG	BL	CG	EG	SL
讀取	1.0 至 2 V	0.6 至 2 V	0 至 2.6 V	0 至 2.6 V	0 V
抹除	-0.5 V/0 V	0 V	0 V/-8 V	8 至 12 V	0 V
程式化	1 V	1 μ A	8 至 11 V	4.5 至 9 V	4.5 至 5 V

【0046】 圖 4 描繪分離閘三閘極記憶體單元 410。記憶體單元 410 與圖 3 的記憶體單元 310 相同，除了記憶體單元 410 不具有分開控制閘。除了沒有控制閘偏壓以外，抹除操作（通過抹除閘抹除）與讀取操作與圖 3 所示者類似。程式化操作的完成也未使用控制閘偏壓，因此為了彌補對於控制閘偏壓的缺乏，源極線上的程式化電壓比較高。

【0047】 表 3 描繪可施加至記憶體單元 410 的端子之用於執行讀取、抹除、及程式化操作的典型電壓範圍：

表 3：圖 4 之快閃記憶體單元 410 的操作

	WL/SG	BL	EG	SL
讀取	0.7 至 2.2 V	0.6 至 2 V	0 至 2.6 V	0 V
抹除	-0.5 V/0 V	0 V	11.5 V	0 V
程式化	1 V	2 至 3 μ A	4.5 V	7 至 9 V

【0048】 圖 5 描繪堆疊閘極記憶體單元 510。記憶體單元 510 類似於圖 2 的記憶體單元 210，除了浮閘 20 在整個通道區域 18 上方延伸，及控制閘 22 在浮閘 20 上方延伸，藉由絕緣層分隔。抹除、程式化、及讀取操作以與先前針對記憶體單元 210 描述的類似方式操作。

【0049】 表 4 描繪可施加至記憶體單元 510 的端子之用於執行讀取、抹除、及程式化操作的典型電壓範圍：

表 4：圖 5 之快閃記憶體單元 510 的操作

	CG	BL	SL	P-sub
讀取	2 至 5 V	0.6 至 2 V	0 V	0 V
抹除	-8 至-10 V/0 V	FLT	FLT	8 至 10 V / 15 至 20 V
程式化	8 至 12 V	3 至 5 V	0 V	0 V

【0050】 為了在人工類神經網路中利用包含上述非揮發性記憶體單元之類型的一者的記憶體陣列，進行二種修改。第一，該等線經組態使得各記憶體單元可個別地程式化、抹除、及讀取而不會不利地影響陣列中之其他記憶體單元的記憶體狀態，如下文所進一步解釋者。第二，提供記憶體單元連續（類比）程式化。

【0051】 具體而言，陣列中的各記憶體單元的記憶體狀態（亦即，浮閘上的電荷）可獨立地且對其他記憶體單元的干擾最小地從完全抹除狀態

連續地改變至完全程式化狀態。在另一實施例中，陣列中的各記憶體單元的記憶體狀態（亦即，浮閘上的電荷）可獨立地且對其他記憶體單元的干擾最小地從完全程式化狀態連續地改變至完全抹除狀態，且反之亦然。此意指單元儲存係類比的或至少可儲存許多離散值（諸如 16 或 64 個不同的值）之一者，其允許非常精確及個別的調諧記憶體陣列中的所有單元，且其使記憶體陣列非常適於儲存類神經網路的突觸權重並對該等突觸權重進行微調調整。

採用非揮發性記憶體單元陣列的類神經網路

【0052】 圖 6 概念性地繪示利用非揮發性記憶體陣列之類神經網路的非限制性實例。此實例將非揮發性記憶體陣列類神經網路使用於面部識別應用，但任何其他適當應用可使用基於非揮發性記憶體陣列的類神經網路實作。

【0053】 S0 係輸入，針對此實例，其係具有 5 位元精確度之 32×32 像素的 RGB 影像（亦即，三個 32×32 像素陣列，一者用於各顏色 R、G、及 B，各像素係 5 位元精確度）。從 S0 到 C1 的突觸 CB1 具有二組不同的權重及共享權重，並以 3×3 像素重疊濾波器（核心），將濾波器移位 1 個像素（或如模型所決定而多於 1 個像素）掃描輸入影像。具體而言，將影像之 3×3 部分中的 9 個像素（亦即，稱為濾波器或核心）的值提供至突觸 CB1，藉此將此等 9 個輸入值乘以合適權重，且在對該乘法的輸出加總之後，由 CB1 的第一神經元判定並提供用於產生特徵圖譜(feature map) C1 之該等層的一者之像素的單一輸出值。然後將 3×3 濾波器右移一個像素（亦

即，在右邊加入三個像素之行，並在左邊丟棄三個像素之行），藉此將此新定位濾波器中的 9 個像素值提供至突觸 CB1，藉此將彼等乘以相同權重，並藉由關聯神經元判定第二單一輸出值。繼續此程序直到 3×3 濾波器針對所有三種顏色及所有位元（精確值）掃描整個 32×32 像素影像。然後使用不同組的權重重複該程序，以產生 C1 之不同的特徵圖譜，直到層 C1 的所有特徵圖譜已計算。

【0054】 在本實例中，在 C1 處，有各具有 30×30 個像素之 16 個特徵圖譜。各像素係從將輸入與核心相乘所擷取的新特徵像素，且因此各特徵圖譜係二維陣列，且因此在此實例中，突觸 CB1 構成二維陣列的 16 層（請記住本文中所提及的神經元層及陣列係邏輯關係，不一定係實體關係-亦即，陣列不一定以實體的二維陣列定向）。16 個特徵圖譜之各者係由經施加至濾波器掃描之十六組不同的突觸權重之一者產生。C1 特徵圖譜可全部關於相同影像特徵的不同態樣，諸如邊界識別。例如，第一圖譜（使用第一組權重產生，用於所有掃描所共享而用以產生此第一圖譜）可識別圓形邊緣，第二圖譜（使用與第一組權重不同的第二組權重產生）可識別矩形邊緣，或某些特徵的縱橫比，並依此類推。

【0055】 啟動功能 P1（彙整）係在從 C1 到 S1 之前施加，其彙整來自各特徵圖譜中之連續、非重疊 2×2 區域的值。彙整階段的目的係平均附近位置（或也能使用 max 函數），以減少，例如，邊緣位置的相依性，並在進入下一階段之前減少資料大小。在 S1 處，有 16 個 15×15 特徵圖譜（亦即，各 15×15 像素的十六個不同陣列）。從 S1 到 C2 之 CB2 中的突觸及關聯神經元使用 4×4 濾波器，以 1 個像素的濾波器移位來掃描 S1 中的圖

譜。在 C2 處，有 22 個 12×12 特徵圖譜。啟動功能 P2（彙整）係在從 C2 到 S2 之前施加，其彙整來自各特徵圖譜中之連續、非重疊 2×2 區域的值。在 S2 處，有 22 個 6×6 特徵圖譜。將啟動功能施加在從 S2 到 C3 的突觸 CB3 處，其中 C3 中的每個神經元連接至 S2 中的每個圖譜。在 C3 處，有 64 個神經元。從 C3 到輸出 S3 的突觸 CB4 將 S3 完全連接至 C3。在 S3 處的輸出包括 10 個神經元，其中最高的輸出神經元判定類別。該輸出可例如指示原始影像之內容的識別或類別。

【0056】 突觸的各級係使用非揮發性記憶體單元的陣列或陣列的一部分實作。圖 7 係包括非揮發性記憶體單元且用作輸入層與次一層之間的突觸之向量矩陣乘法(vector-by-matrix multiplication, VMM)陣列的方塊圖。具體而言，VMM 32 包括非揮發性記憶體單元 33 的陣列、抹除閘及字線閘解碼器 34、控制閘解碼器 35、位元線解碼器 36、及源極線解碼器 37，其等解碼記憶體陣列 33 的輸入。此實例中的源極線解碼器 37 也解碼記憶體單元陣列的輸出。替代地，位元線解碼器 36 可解碼記憶體陣列的輸出。記憶體陣列用作兩個目的。第一，其儲存將由 VMM 使用的權重。第二，記憶體陣列有效地將輸入乘以儲存在記憶體陣列中的權重並將其等每輸出線（源極線或位元線）加總以產生輸出，其將係至次一層的輸入或至最終層的輸入。藉由執行乘法及加法功能，記憶體陣列否定對分開的乘法及加法邏輯電路的需求，且亦由於原位記憶體運算而係有能源效率的。

【0057】 將記憶體陣列的輸出供應至一差動加算器（諸如加總運算放大器）38，其加總記憶體單元陣列的輸出以創建用於該卷積的單一值。該差動加算器係諸如用以實現正權重及負權重與正輸入的加總。然後將該經

加總輸出值供應至整流輸出的啟動功能電路 39。啟動功能可包括 S 形 (sigmoid) 函數、雙曲正切(tanh)函數、ReLU 函數。經整流輸出值變成作為次一層（例如，以上描述中的 C1）之特徵圖譜的元素，然後經施加於次一突觸以產生次一特徵圖譜層或最終層。因此，在此實例中，記憶體陣列構成複數個突觸（其從神經元的先前層或從輸入層（諸如影像資料庫）接收其等的輸入），且加總運算放大器 38 及啟動功能電路 39 構成複數個神經元。

【0058】 圖 8 係 VMM 之各種級的方塊圖。輸入係藉由數位轉類比轉換器 31 從數位轉換成類比，並提供至輸入 VMM 32a。將由輸入 VMM 32a 產生的輸出提供作為至次一 VMM（隱藏級 1）32b 的輸入，其繼而產生提供作為至次一 VMM（隱藏級 2）32b 之輸入的輸出，並依此類推。VMM 32 之各種層如卷積式類神經網路(convolutional neural network, CNN)之突觸及神經元的不同層而作用。各 VMM 可係獨立的非揮發性記憶體陣列，或多個 VMM 可利用相同的非揮發性記憶體陣列的不同部分，或多個 VMM 可利用相同的非揮發性記憶體陣列的重疊部分。

向量矩陣乘法(VMM)陣列

【0059】 圖 9 描繪神經元 VMM 900，其特別適於圖 2 所示之類型的記憶體單元，且利用為輸入層與次一層之間的神經元的突觸及部件。VMM 900 包含非揮發性記憶體單元的記憶體陣列 903、參考陣列 901、及參考陣列 902。參考陣列 901 及參考陣列 902 用以將流入端子 BLR0 至 3 中的電流輸入轉換成電壓輸入 WL0 至 3。如圖示之參考陣列 901 及 902 係在行方向

上。一般而言，參考陣列方向正交於輸入線。實際上，參考記憶體單元係通過多工器（多工器 914，其包括多工器及用於偏壓參考位元線的疊接電晶體 VBLR）的二極體式連接，該等多工器具有流入其等的電流輸入。參考單元經調諧至目標參考位準。

【0060】 記憶體陣列 903 供兩個目的使用。第一，其儲存將由 VMM 900 使用的權重。第二，記憶體陣列 903 有效地將輸入（在端子 BLR0 至 3 中提供的電流輸入；參考陣列 901 及 902 將此等電流輸入轉換成輸入電壓以供應至字線 WL0 至 3）乘以儲存在記憶體陣列中的權重以產生輸出，該輸出將係至次一層的輸入或至最終層的輸入。藉由實施乘法功能，記憶體陣列否定對分開的乘法邏輯電路的需求，且也係有能源效率的。此處，電壓輸入係在字線上提供，且輸出在一讀取（推論）操作期間顯現於位元線上。置於位元線上的電流執行來自連接至位元線之記憶體單元的所有電流之一加總功能。

【0061】 圖 10 描繪用於 VMM 900 的操作電壓。表中的行指示置於用於經選擇單元之字線、用於未經選擇單元之字線、用於經選擇單元之位元線、用於未經選擇單元之位元線、用於經選擇單元之源極線、以及用於未經選擇單元之源極線上的電壓。列指示讀取、抹除、及程式化操作。

【0062】 圖 11 描繪神經元 VMM 1100，其特別適於圖 2 所示之類型的記憶體單元，且利用為輸入層與次一層之間的神經元的突觸及部件。VMM 1100 包含非揮發性記憶體單元的記憶體陣列 1103、參考陣列 1101、及參考陣列 1102。VMM 1100 類似於 VMM 900，除了在 VMM 1100 中，字線在垂直方向上延行。有兩個參考陣列 1101（在頂部，其提供將輸入電

流轉換成用於偶數列之電壓的參考)及 1102 (在底部,其提供將輸入電流轉換成用於奇數列之電壓的參考)。此處,輸入係在字線上提供,且輸出在一讀取操作期間顯現於源極線上。置於源極線上的電流執行來自連接至源極線之記憶體單元的所有電流之一加總功能。

【0063】 圖 12 描繪用於 VMM 1100 的操作電壓。表中的行指示置於用於經選擇單元之字線、用於未經選擇單元之字線、用於經選擇單元之位元線、用於未經選擇單元之位元線、用於經選擇單元之源極線、以及用於未經選擇單元之源極線上的電壓。列指示讀取、抹除、及程式化操作。

【0064】 圖 13 描繪神經元 VMM 1300,其特別適於圖 3 所示之類型的記憶體單元,且利用為輸入層與次一層之間的神經元的突觸及部件。VMM 1300 包含非揮發性記憶體單元之記憶體陣列 1301、參考陣列 1302 (提供將輸入電流轉換成用於偶數列之輸入電壓的參考)、及參考陣列 1303 (其提供將輸入電流轉換成用於奇數列之輸入電壓的參考)。VMM 1300 類似於 VMM 900,除了 VMM 1300 進一步包含耦接至一系列記憶體單元之控制閘的控制線 1306,以及耦接至鄰接列之記憶體單元之抹除閘的控制線 1307。此處,字線、控制閘線、及抹除閘線具有相同的方向。VMM 進一步包含參考位元線選擇電晶體 1304 (多工器 1314 之部分),其選擇性將參考位元線耦接至經選擇參考記憶體單元之位元線接觸件,以及開關 1305 (多工器 1314 之部分),其選擇性將參考位元線耦接至用於特定之經選擇參考記憶體單元的控制線 1306。此處,輸入係在(記憶體陣列 1301 之)字線上提供,且輸出在一讀取操作期間出現於位元線(諸如位元線 1309)上。置於位元線上的電流執行來自連接至位元線之記憶體單元的所

有電流的加總功能。

【0065】 圖 14 描繪用於 VMM 1300 的操作電壓。表中的行指示置於用於經選擇單元之字線、用於未經選擇單元之字線、用於經選擇單元之位元線、用於未經選擇單元之位元線、用於經選擇單元之控制閘、用於與經選擇單元相同磁區中之未經選擇單元之控制閘、用於不同於經選擇單元之一磁區中之未經選擇單元之控制閘、用於經選擇單元之抹除閘、用於未經選擇單元之抹除閘、用於經選擇單元之源極線、用於未經選擇單元之源極線上的電壓。列指示讀取、抹除、及程式化操作。

【0066】 圖 15 描繪神經元 VMM 1500，其特別適於圖 3 所示之類型的記憶體單元，且利用為輸入層與次一層之間的神經元的突觸及部件。VMM 1500 類似於 VMM 1100，除了在 VMM 1500 中，抹除閘線（諸如抹除閘線 1501）在垂直方向上延行。此處，輸入經提供在字線上，且輸出出現在源極線上。置於位元線上的電流執行來自連接至位元線之記憶體單元的所有電流的加總功能。

【0067】 圖 16 描繪用於 VMM 1500 的操作電壓。表中的行指示置於用於經選擇單元之字線、用於未經選擇單元之字線、用於經選擇單元之位元線、用於未經選擇單元之位元線、用於經選擇單元之控制閘、用於與經選擇單元相同磁區中之未經選擇單元之控制閘、用於不同於經選擇單元之一磁區中之未經選擇單元之控制閘、用於經選擇單元之抹除閘、用於未經選擇單元之抹除閘、用於經選擇單元之源極線、用於未經選擇單元之源極線上的電壓。列指示讀取、抹除、及程式化操作。

【0068】 圖 17 描繪神經元 VMM 1700，其特別適於圖 3 所示之類型

的記憶體單元，且利用為輸入層與次一層之間的神經元的突觸及部件。

VMM 1700 包含非揮發性記憶體單元的記憶體陣列 1701 及參考陣列 1702（在陣列的頂部）。替代地，可將另一參考陣列置於底部，類似於圖 10 者。

在其他方面，VMM 1700 類似於先前描述之 VMM（除了在 VMM 1700 中），控制閘線（諸如控制閘線 1703）在垂直方向上延行（因此參考陣列 1702 在列方向上，正交於輸入控制閘線），且抹除閘線（諸如抹除閘線 1704）在水平方向上延行。此處，輸入經提供在控制閘線上，且輸出出現在源極線上。在一個實施例中，僅使用偶數列，且在另一實施例中，僅使用奇數列。置於源極線上的電流執行來自連接至源極線之記憶體單元的所有電流之一加總功能。

【0069】 如本文針對類神經網路所描述的，快閃單元較佳地經組態以在次臨限區域中操作。

【0070】 在本文中描述的記憶體單元以弱反轉偏壓：

$$I_{ds} = I_o * e^{(V_g - V_{th})/kV_t} = w * I_o * e^{(V_g)/kV_t}$$

$$w = e^{(-V_{th})/kV_t}$$

【0071】 對於使用記憶體單元以將輸入電流轉換成輸入電壓的 I 至 V 對數轉換器：

$$V_g = k * V_t * \log [I_{ds}/w * I_o]$$

【0072】 對於使用作為向量矩陣乘法器 VMM 的記憶體陣列，輸出電流係：

$$I_{out} = w_a * I_o * e^{(V_g)/kV_t}, \text{ 即}$$

$$I_{out} = (w_a/w_p) * I_{in} = W * I_{in}$$

$$W = e^{(V_{thp} - V_{tha})/kV_t}$$

【0073】 字線或控制閘可使用為用於輸入電壓的記憶體單元之輸入。

【0074】 替代地，快閃記憶體單元可經組態以在線性區域中操作：

$$I_{ds} = \beta * (V_{gs} - V_{th}) * V_{ds}; \beta = \mu * C_{ox} * W/L$$

$$W \propto (V_{gs} - V_{th})$$

【0075】 對於 I 至 V 線性轉換器，在線性區域中操作的記憶體單元可使用以將輸入/輸出電流線性地轉換成輸入/輸出電壓。

【0076】 ESF 向量矩陣乘法器的其他實施例描述於美國專利申請案第 15/826,345 號中，其以引用方式併入本文中。源極線或位元線可使用為神經元輸出（電流加總輸出）。

【0077】 圖 18 描繪類比神經形態記憶體系統 1800，其包含系統控制器 1801、類比神經形態記憶體引擎 1802、及類比神經形態記憶體引擎 1803。類比神經形態記憶體引擎 1802 及類比神經形態記憶體引擎 1803 各自含有以權重來程式化的 VNM 陣列。類比神經形態記憶體引擎 1803 中的 VNM 中的陣列之一部分專用於用於判定資料是否已漂移的程序。

【0078】 圖 19 描繪用於再新在 VNM 中資料的方法 1900。取樣一組快閃記憶體單元（在步驟 1901 中樣本選擇）。針對由快閃記憶體單元 (L0, ... LN-1) 所儲存的各可能的位準 (N 值) 分析樣本，且在各對相鄰位準之間計算一差值 (DR_x)（步驟 1902）。針對各位準，將差值 (DR_x) 與目

標（理想）(DRx_target)比較（步驟 1903）。若差值(DRx)之任一者超過該目標，則設定一再新旗標（步驟 1904）。然後發生還原資料程序（步驟 1905）。此步驟可藉由晶片上(on-chip)控制器或藉由晶片外(off-chip)控制器來完成。在步驟 1905，再調諧步驟（例如，再程式化）係針對各位準所定製。首先，單元經讀回，然後取決於該位準，對該位準施加差量再調諧。針對此目的可使用查找表（Icell 對 I-tuned target）。在此步驟中，使用調諧演算法（疊代驗證及程式化）以達到該目標（諸如，藉由使用美國專利申請案第 15/826,345 號中所述的演算法，其以引用方式併入本文中）。對於所取樣之資料保留單元，回復亦經完成。一旦各單元具有等於儲存在該單元中的位準之目標電流（步驟 1906）的電流，則程序完成（步驟 1908）。若從未達到彼情況，則單元被認定為不良且程序完成（步驟 1907）。

【0079】 圖 20 描繪用於再新在 VNM 中資料的另一方法 2000。取樣一組例示性快閃記憶體單元（步驟 2001）。針對由快閃記憶體單元($L_0, \dots, L_{N-1}$)所儲存的各可能的位準（N 值）分析樣本，且在各對相鄰位準之間計算差值(DRx)（步驟 2002）。針對各位準，將差值(DRx)與目標（理想）(DRx_target)比較（步驟 2003）。若差值(DRx)之任一者超過該目標，則設定一再新旗標（步驟 2004）。然後發生還原資料程序（步驟 2005）。可藉由晶片上或晶片外控制器來完成該還原。在步驟 2005，再調諧步驟（例如，再程式化）係針對各位準所定製。首先，單元（陣列單元及經取樣之資料保留單元）經讀回，然後取決於該位準，藉由微調增量還原演算法對該位準施加差量再調諧。針對此目的可使用查找表（Icell 對 I-tuned target）。在此步驟中，使用微調調諧演算法（微調步驟疊代驗證及程式化）

以達到該目標（諸如，藉由使用美國專利申請案第 15/826,345 號中所述的演算法，其以引用方式併入本文中）。替代地，可藉由將來自訓練的所有權重轉移至 VMM 陣列而再程式化整個 VMM 陣列來完成資料還原。一旦各單元具有等於儲存在該單元中的位準之目標電流（步驟 2006）的電流，則程序完成（步驟 2007）。若從未達到彼情況，則單元被認定為不良（步驟 2008）。然後使用冗餘陣列磁區置換不良陣列磁區（步驟 2009）。若以此方式所使用的冗餘單元之數目超過臨限（步驟 2010），則整個磁區被認定為不良（步驟 2011）。

【0080】 關於圖 21 及圖 22，可係所欲的是識別「快速單元」，其等係在程式化操作期間相較於平均單元可從其等之浮閘相對快速地失去電荷的單元，因為此等快速單元隨著時間更易於發生資料漂移。在圖 21 中，可參照圖表 2100 來識別快速單元，如以相對較少的程式化嘗試（計數）而獲得高電流位準(Ircell)者。在圖 22 中，參照圖表 2200，應理解的是，快速單元較正常單元可能具有高 Ir 操作範圍。圖 22 顯示用以資料保持（漂移）監測器的例示性單元。超過最小及最大範圍的單元係用於監測目的。此係用以創建具有較正常單元差的應力狀況的監測器單元。

【0081】 圖 23 描繪資料漂移偵測方案。在此實例中，在目標 1 與目標 2 之間存在著特定電流位準係所欲的。若所感測之電流介於該等目標之間，則不需要再新，此意味著對於系統性能而言，漂移係可接受的。然而，若所感測之電流係在目標 1 與目標 2 之間的範圍外，則執行一資料再新（還原）程序。

【0082】 圖 24 描繪資料漂移偵測器之實施例。資料漂移偵測器 2400

接收電流 2402（其可係通過經選擇單元的感測電流位準或此等電流之間的差值）以及比較器 2403 中的參考電流 2401，其輸出指示電流 2402 是否超過參考電流 2401 的值。資料漂移偵測器 2400 基於意欲已儲存在經選擇單元中的值，比較由經選擇單元所汲取的電流與目標電流。

【0083】 圖 25 描繪資料漂移偵測器之另一實施例。資料漂移偵測器 2500 包含取樣及保持電容器 2501 與 2504、開關 2502 及 2505、電流源 2506（其可係通過經選擇單元的感測電流位準或此等電流之間的差值）、參考電流 2503、及比較器 2507。在此實施例中，兩個電流源係由該等電容器通過經判定之斜坡時間時期取樣並保持，然後由比較器 2507 比較該等經取樣之值。因此，資料漂移偵測器 2500 基於意欲已儲存在經選擇單元中的值，比較由經選擇單元所汲取的經取樣電流與經取樣目標電流。

【0084】 參照圖 26，如圖表 2600 中所顯示，可基於對在發生顯著的資料漂移前可能有多少時間經過之瞭解來選擇資料漂移偵測器 2500 之取樣時期。相較於資料漂移偵測器 2400，此可導致較少的功率消耗。

【0085】 圖 27 描繪資料漂移偵測器之另一實施例。資料漂移偵測器 2700 包含經選擇之記憶體單元 2701、電流源 2702、參考電壓 2703、及比較器 2704。在此實施例中，在經選擇之記憶體單元 2701 之浮閘上儲存的電壓（節點 2705 上的電壓 $\approx V_{FG} - V_{tFG}$ ）與參考電壓 2703（其係經選擇之記憶體單元 2701 意欲儲存的電壓）之間進行直接比較。記憶體單元 2701 係以一源極隨耦器操作模式組態。參考電壓可係自一參考記憶體單元所產生。

【0086】 應注意的是，如本文中所使用，「在...上方(over)」及

「在...上(on)」之用語皆含括性地包括了「直接在...之上(directly on)」(無居中的材料、元件或間隔設置於其間)及「間接在...之上(indirectly on)」(有居中的材料、元件或間隔設置於其間)的含意。同樣地，「相鄰的(adjacent)」一詞包括了「直接相鄰的」(無居中的材料、元件或間隔設置於其間)及「間接相鄰的」(有居中的材料、元件或間隔設置於其間)的含意，「安裝於(mounted to)」一詞則包括了「直接安裝於(directly mounted to)」(無居中的材料、元件或間隔設置於其間)及「間接安裝於(indirectly mounted to)」(有居中的材料、元件或間隔設置於其間)的含意，以及「電耦接(electrically coupled)」一詞則包括了「直接電耦接(directly electrically coupled to)」(無居中的材料或元件於其間將各元件電性相連接)及「間接電耦接(indirectly electrically coupled to)」(有居中的材料或元件於其間將各元件電性相連接)的含意。舉例而言，「在基材上方(over a substrate)」形成元件可包括直接在基材上形成元件而其間無居中的材料/元件存在，以及間接在基材上形成元件而其間有一或多個居中的材料/元件存在。

【符號說明】

【0087】

12...半導體基材/基材

14...源極區域/汲極

16...汲極區域/源極/源極區域/汲極

18...通道區域

20...浮閘

22...字線端子/控制閘

24...位元線
 26...閘極氧化物
 28...選擇閘
 30...抹除閘
 31...數位轉類比轉換器
 32...VMM
 32a...VMM
 32b...VMM（隱藏級 1）/ VMM（隱藏級 2）
 33...非揮發性記憶體單元/記憶體陣列
 34...抹除閘及字線閘解碼器
 35...控制閘解碼器
 36...位元線解碼器
 37...源極線解碼器
 38...差動加算器/加總運算放大器
 39...啟動功能電路
 210...記憶體單元/快閃記憶體單元
 310...四閘極記憶體單元/快閃記憶體單元/記憶體單元
 410...分離閘三閘極記憶體單元/快閃記憶體單元/記憶體單元
 510...堆疊閘極記憶體單元/快閃記憶體單元/記憶體單元
 900...VMM
 901...參考陣列
 902...參考陣列
 903...記憶體陣列
 914...多工器
 1100...VMM
 1101...參考陣列
 1102...參考陣列
 1103...記憶體陣列
 1300...VMM

- 1301...記憶體陣列
- 1302...參考陣列
- 1303...參考陣列
- 1304...參考位元線選擇電晶體
- 1305...開關
- 1306...控制線
- 1307...控制線
- 1309...位元線
- 1314...多工器
- 1500...VMM
- 1501...抹除閘線
- 1700...VMM
- 1701...記憶體陣列
- 1702...參考陣列
- 1703...控制閘線
- 1704...抹除閘線
- 1800...類比神經形態記憶體系統
- 1801...系統控制器
- 1802...類比神經形態記憶體引擎
- 1803...類比神經形態記憶體引擎
- 1900...方法
- 1901...步驟
- 1902...步驟
- 1903...步驟
- 1904...步驟
- 1905...步驟
- 1906...步驟
- 1907...步驟
- 1908...步驟

- 2000... 方法
- 2001... 步驟
- 2002... 步驟
- 2003... 步驟
- 2004... 步驟
- 2005... 步驟
- 2006... 步驟
- 2007... 步驟
- 2008... 步驟
- 2009... 步驟
- 2010... 步驟
- 2011... 步驟
- 2100... 圖表
- 2200... 圖表
- 2400... 資料漂移偵測器
 - 2401... 參考電流
 - 2402... 電流
 - 2403... 比較器
- 2500... 資料漂移偵測器
 - 2501... 取樣及保持電容器
 - 2502... 開關
 - 2503... 參考電流
 - 2504... 取樣及保持電容器
 - 2505... 開關
 - 2506... 電流源
 - 2507... 比較器
- 2600... 圖表
- 2700... 資料漂移偵測器
 - 2701... 記憶體單元

- 2702... 電流源
- 2703... 參考電壓
- 2704... 比較器
- 2705... 節點
- BLR0... 流入端子/端子
- BLR1... 流入端子/端子
- BLR2... 流入端子/端子
- BLR3... 流入端子/端子
- CB1... 突觸
- CB2... 突觸
- CB3... 突觸
- CB4... 突觸
- C1... 特徵圖譜/層
- P1... 啟動功能
- P2... 啟動功能
- VBLR... 疊接電晶體
- WL0... 電壓輸入/字線
- WL1... 電壓輸入/字線
- WL2... 電壓輸入/字線
- WL3... 電壓輸入/字線

【發明申請專利範圍】

【請求項1】 一種在類比神經形態記憶體系統中的快速單元中偵測資料漂移的方法，該方法包含：

識別一非揮發性記憶體單元陣列中之快速單元；

取樣該等快速單元以用於資料漂移偵測；

程式化該等快速單元至一目標值以用於資料漂移偵測；及

監測經取樣之該等快速單元以用於資料漂移偵測。

【請求項2】 如請求項 1 之方法，其中該監測步驟包含：

感測在一讀取操作期間由一經取樣之快速單元所汲取的一第一電流；

比較該第一電流與由一參考電路所汲取的一第二電流；

比較該第一電流與由該參考電路所汲取的一第三電流；以及

當該第一電流超過該第三電流時，再新儲存在該經取樣之快速單元之一浮閘中的資料直到該經取樣之快速單元在一讀取操作期間汲取介於該第二電流與該第三電流之間的一經還原電流，以及當該第一電流超過該第二電流且不超過該第三電流時，避免再新該資料。

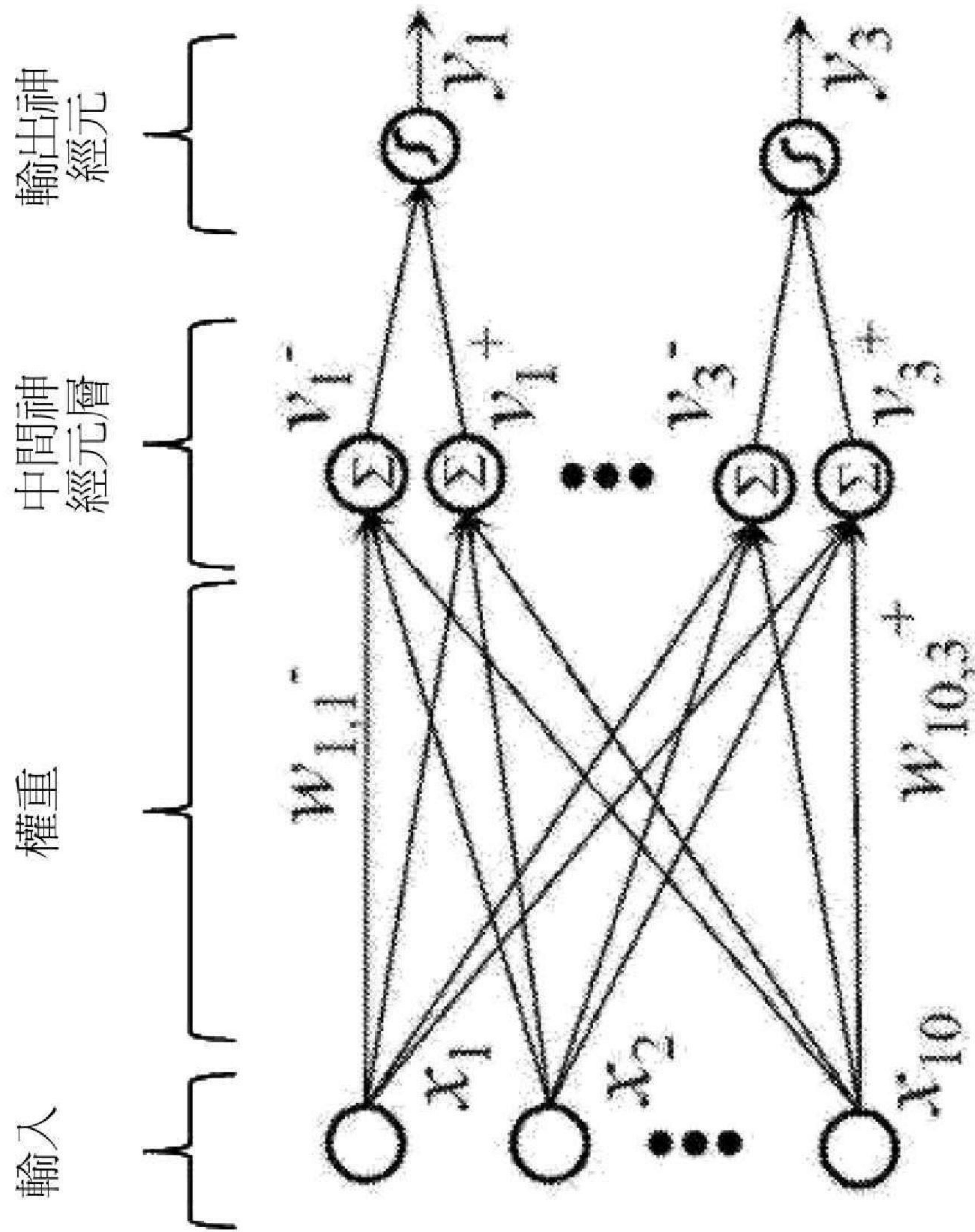
【請求項3】 如請求項 2 之方法，其中該參考電路包含一電流源。

【請求項4】 如請求項 2 之方法，其中該再新步驟係以與一電流位準相關聯的一還原量執行。

【請求項5】 如請求項 2 之方法，其中該再新步驟包含再轉移一訓練權重至該經取樣之快速單元。

【請求項6】 如請求項 1 至 5 中任一項之方法，其中經取樣之該等快速單元係分離閘快閃記憶體單元。

【請求項7】 如請求項 1 至 5 中任一項之方法，其中經取樣之該等快速
單元係堆疊閘極快閃記憶體單元。

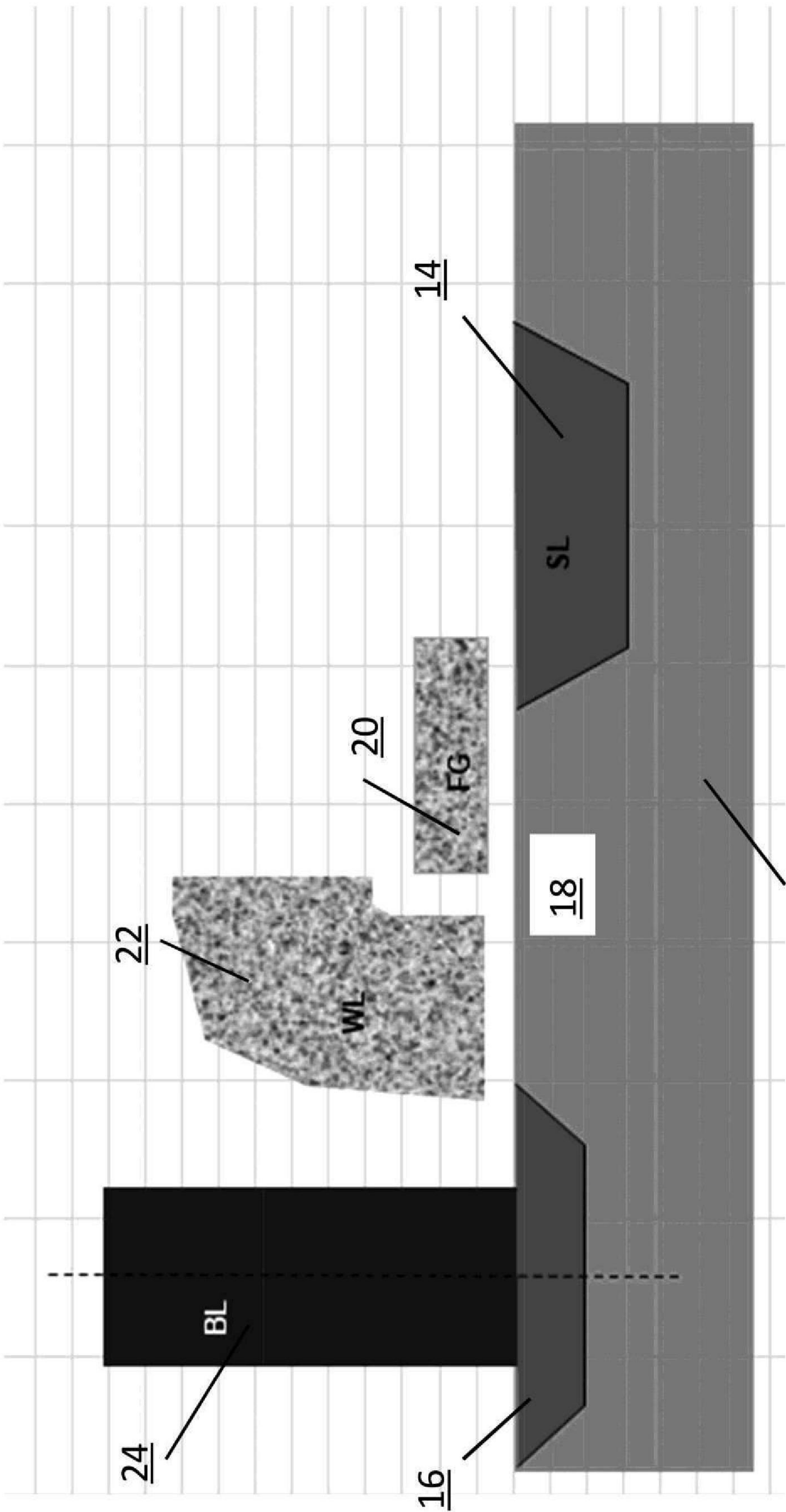


(先前技術)

【圖1】

(先前技術)

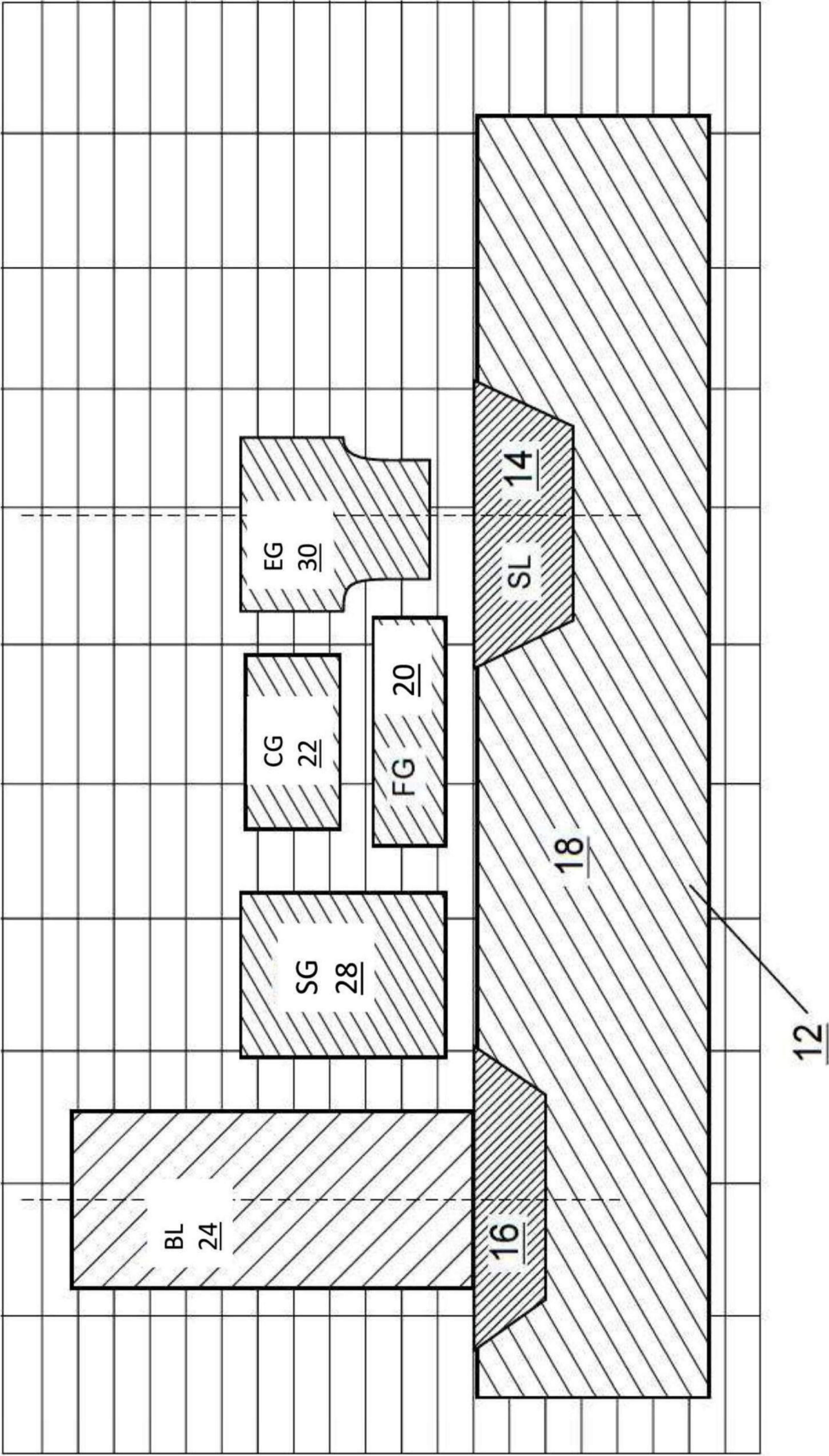
210



【圖2】

(先前技術)

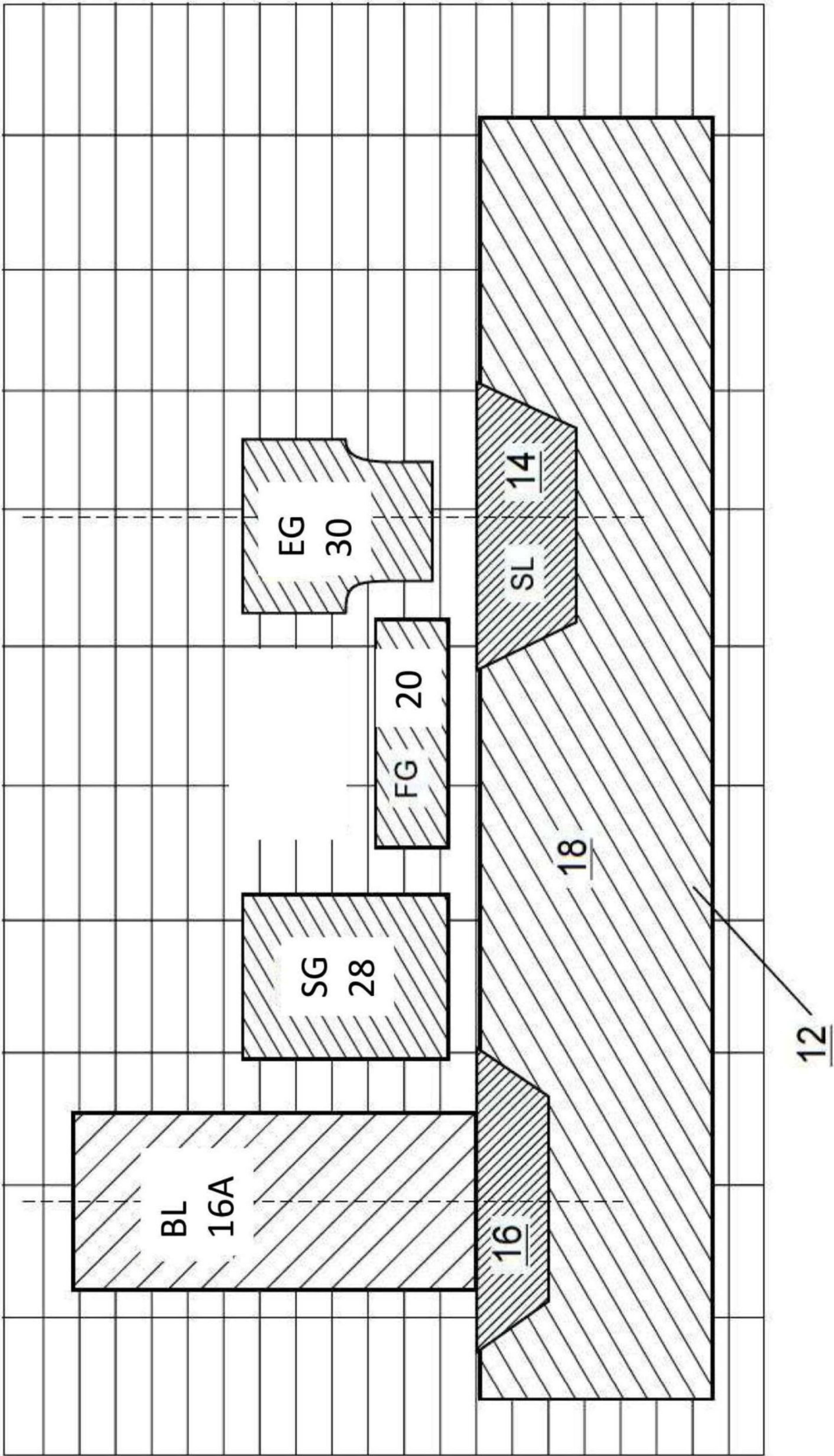
310



【圖3】

(先前技術)

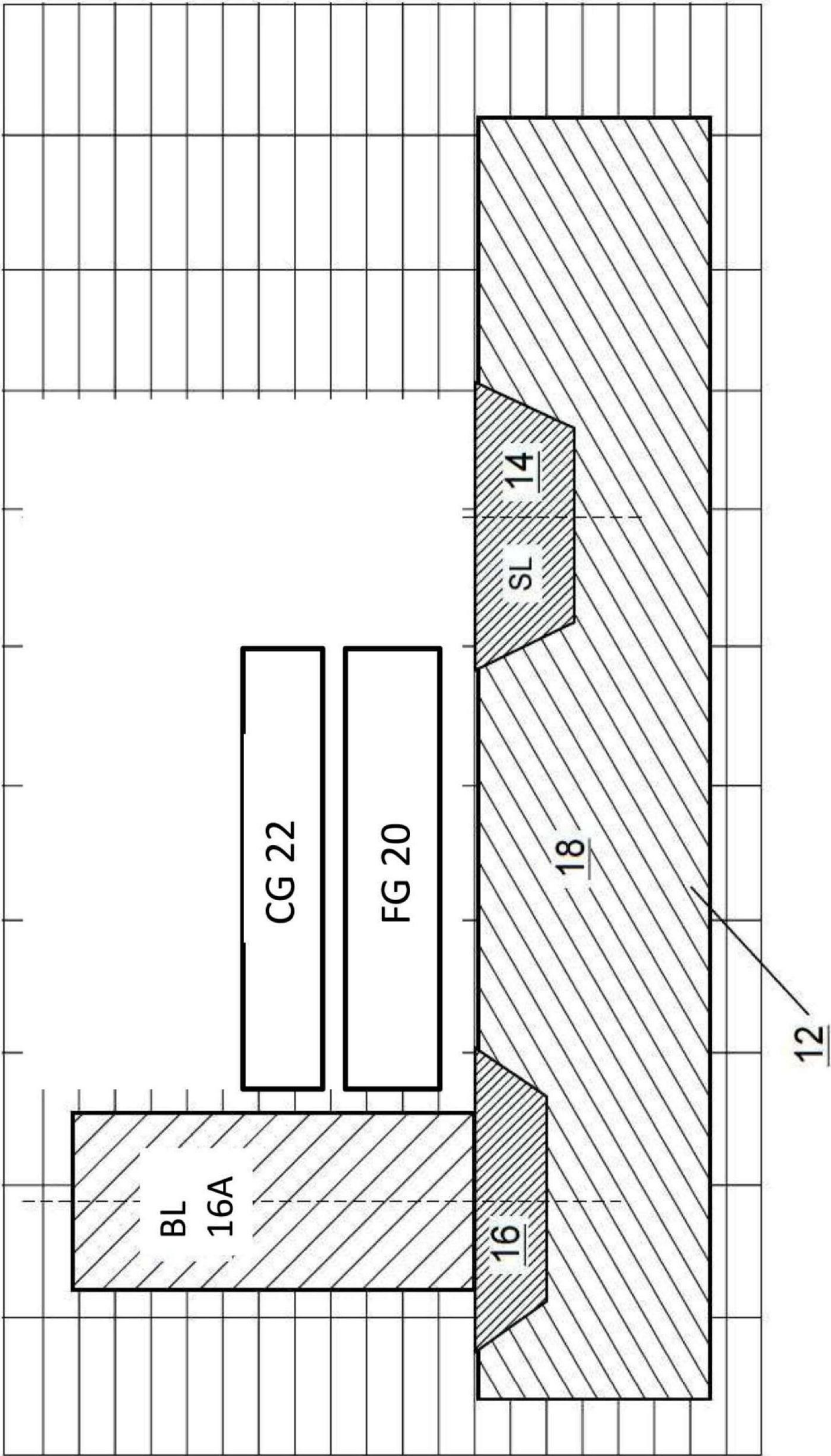
410



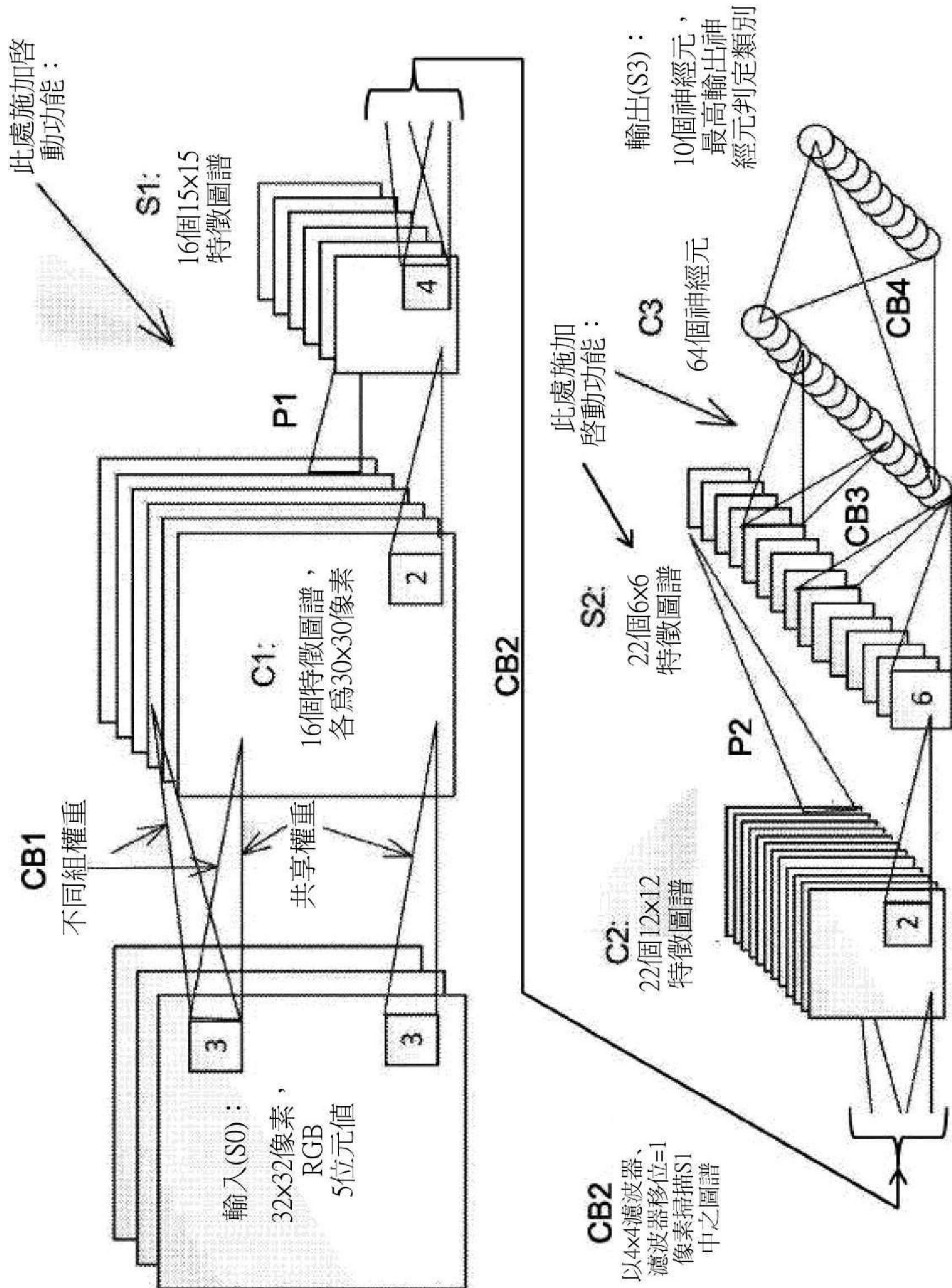
【圖4】

(先前技術)

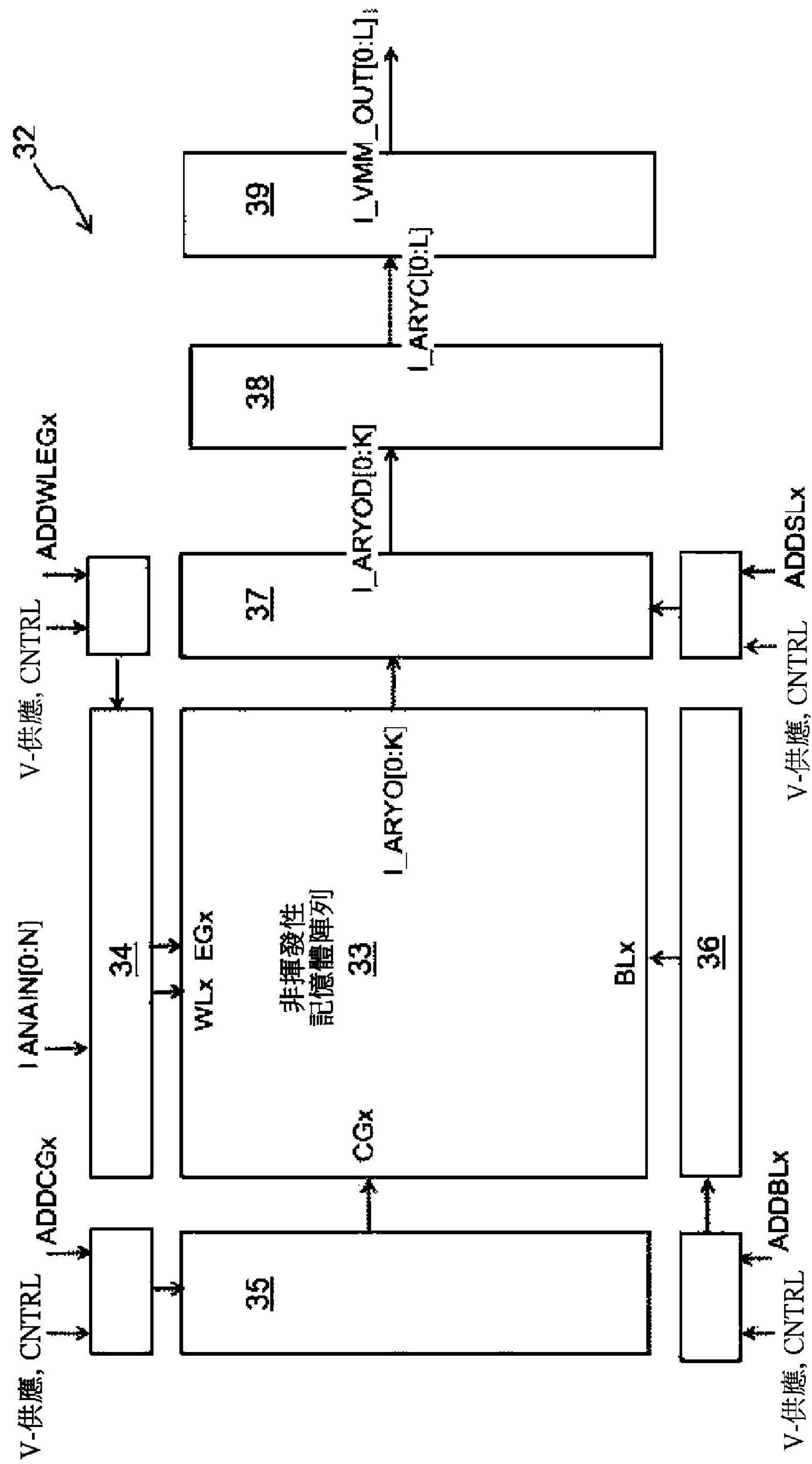
510



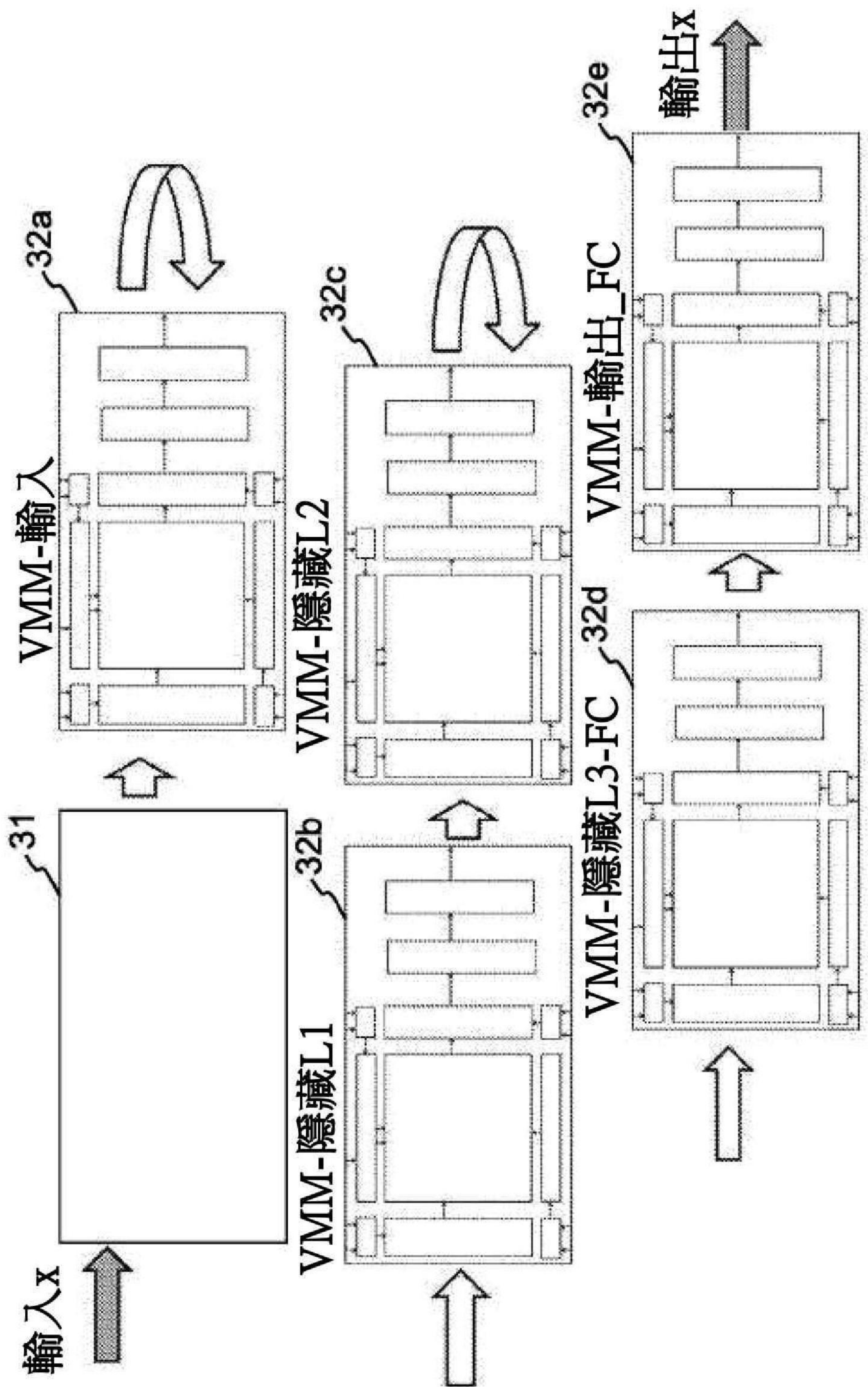
【圖5】



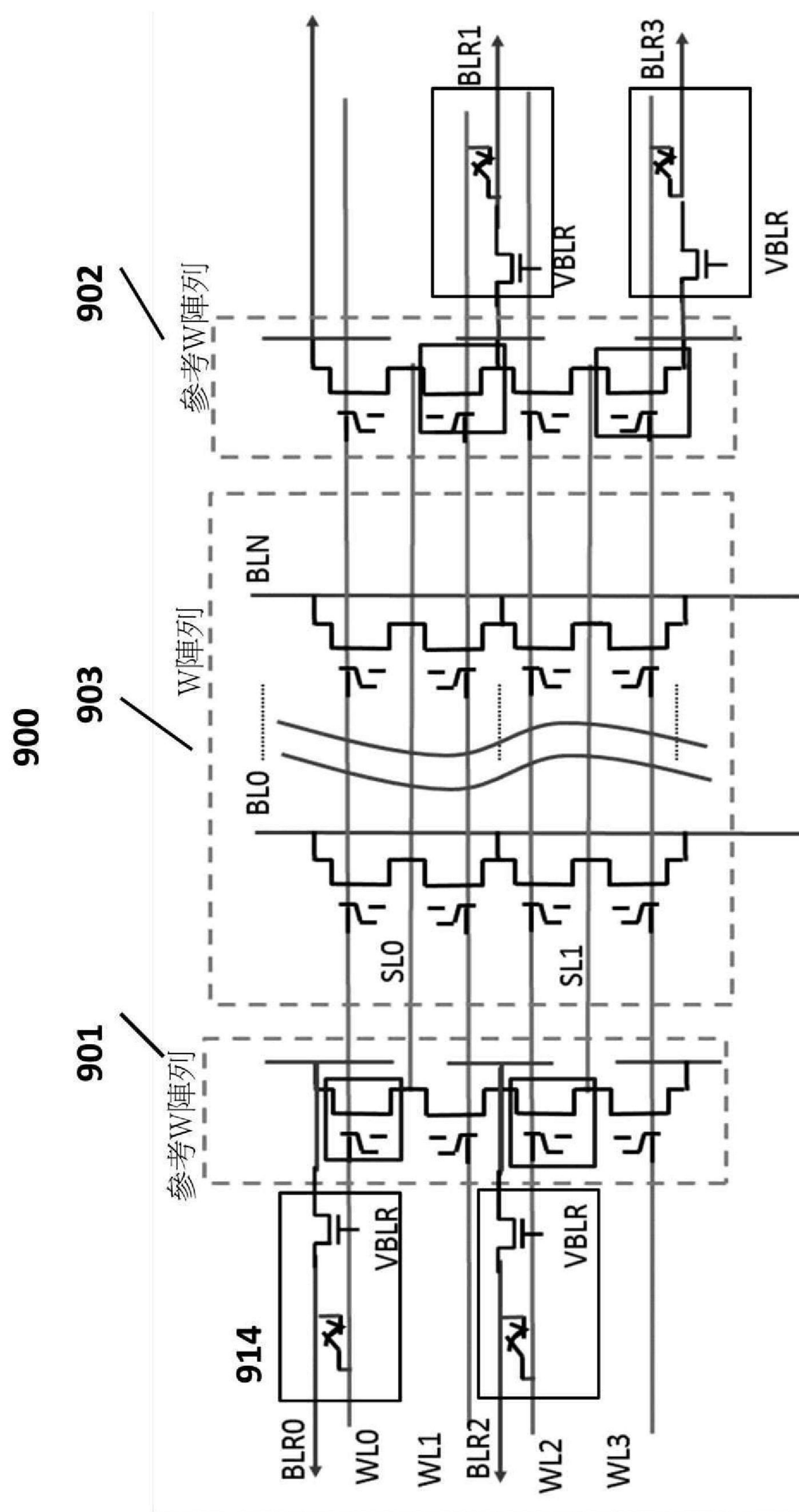
【圖6】



【圖7】



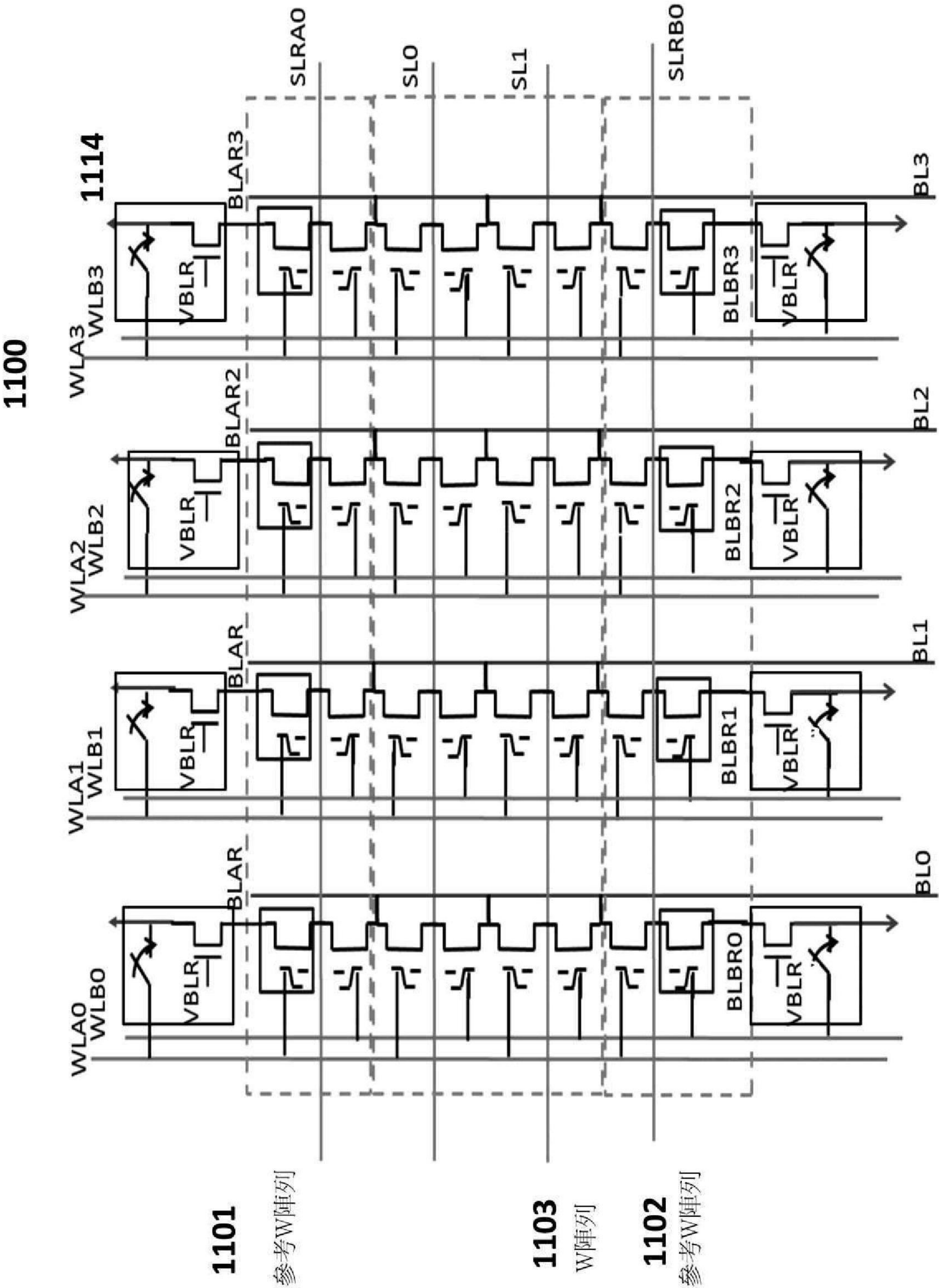
【圖8】



【回】

	WL	WL -未經選擇	BL	BL -未經選擇	SL	SL -未經選擇
讀取	1至3.5V	-0.5 V/0 V	0.6至2 V (Inuron)	0.6V至2V/0V	0 V	0 V
抹除	~5至13V	0 V	0 V	0 V	0 V	0 V
程式化	1至2V	-0.5 V/0 V	0.1至3uA	Vinh ~2.5 V	4至10V	0至1 V/FLT

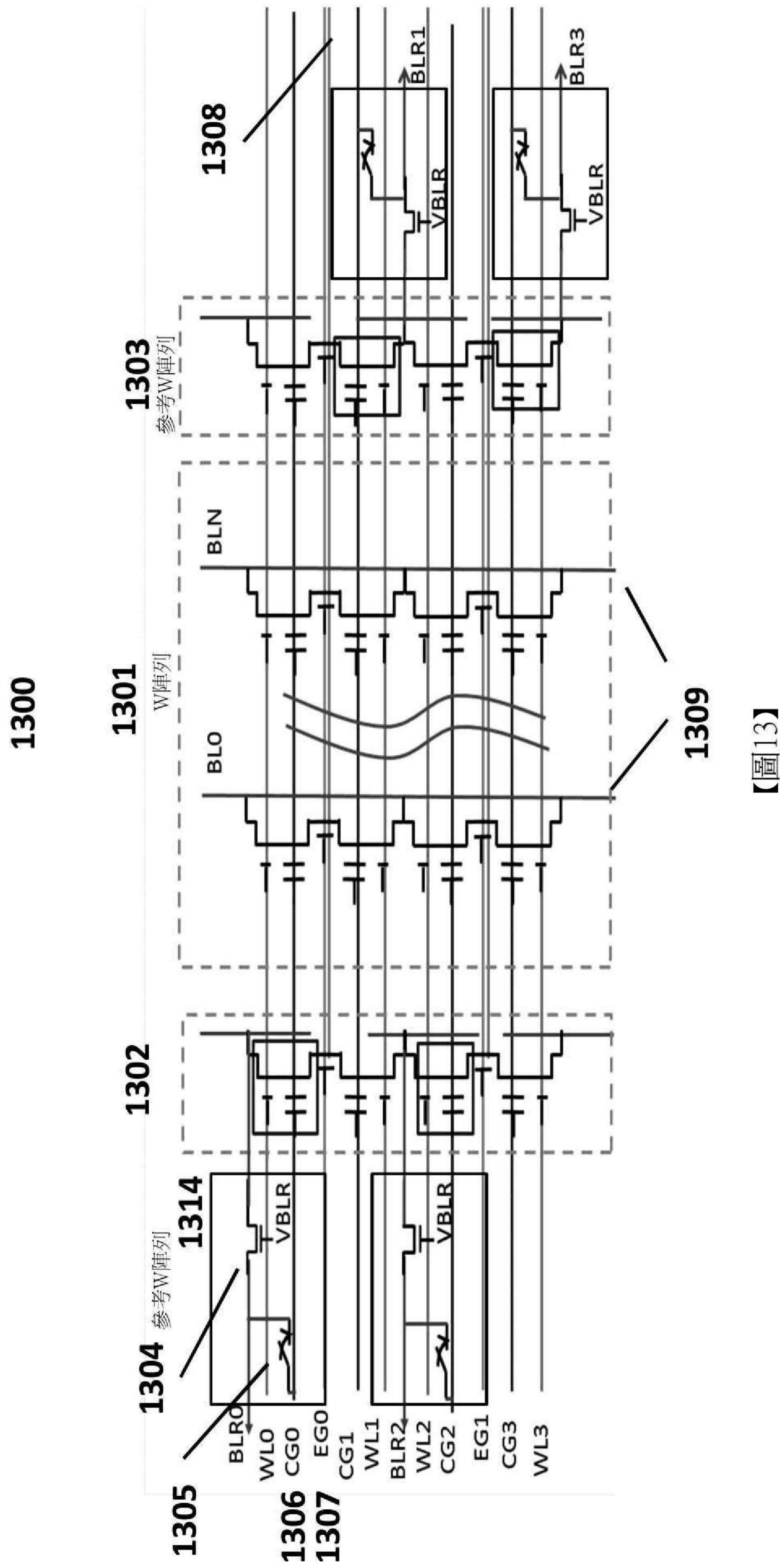
【圖10】



【圖11】

	WL	WL - 未經選擇	BL	BL - 未經選擇	SL	SL - 未經選擇
讀取	1至3.5V	-0.5 V/0 V	0.6至2V	0.6V至2V/0V	~1 V (Ineuron)	0 V
抹除	~5至13V	0 V	0 V	0 V	0 V	SL-inhibit (~4至8 V)
程式化	1至2V	-0.5 V/0 V	0.1至3 uA	Vinh ~2.5 V	4至10V	0至1 V/FLT

【圖12】



【圖13】

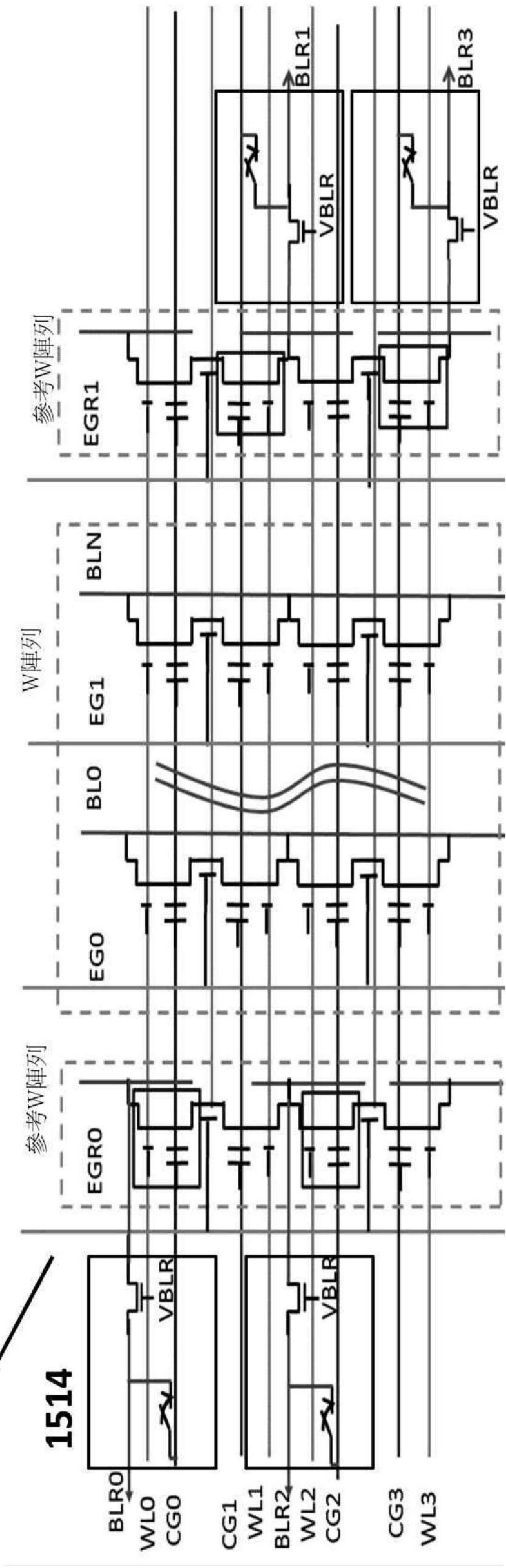
	WL	WL- 未經選擇	BL	BL- 未經選擇	CG	CG-未經選擇 擇相同磁區	CG- 未經選擇	EG	EG- 未經選擇	SL	SL- 未經選擇
讀取	1.0至2V	-0.5 V/ 0 V	0.6至2 V (Inuron)	0 V	0至2.6V	0至2.6V	0至2.6V	0至2.6V	0至2.6V	0 V	0 V
抹除	0 V	0 V	0 V	0 V	0 V	0至2.6V	0至2.6V	5至12V	0至2.6V	0 V	0 V
程式化	0.7至1V	-0.5 V/ 0 V	0.1至1 uA	Vinh 1至2 V)	4至11V	0至2.6V	0至2.6V	4.5至5V	0至2.6V	4.5至5V	0至1V

【圖14】

1500

1501

1514



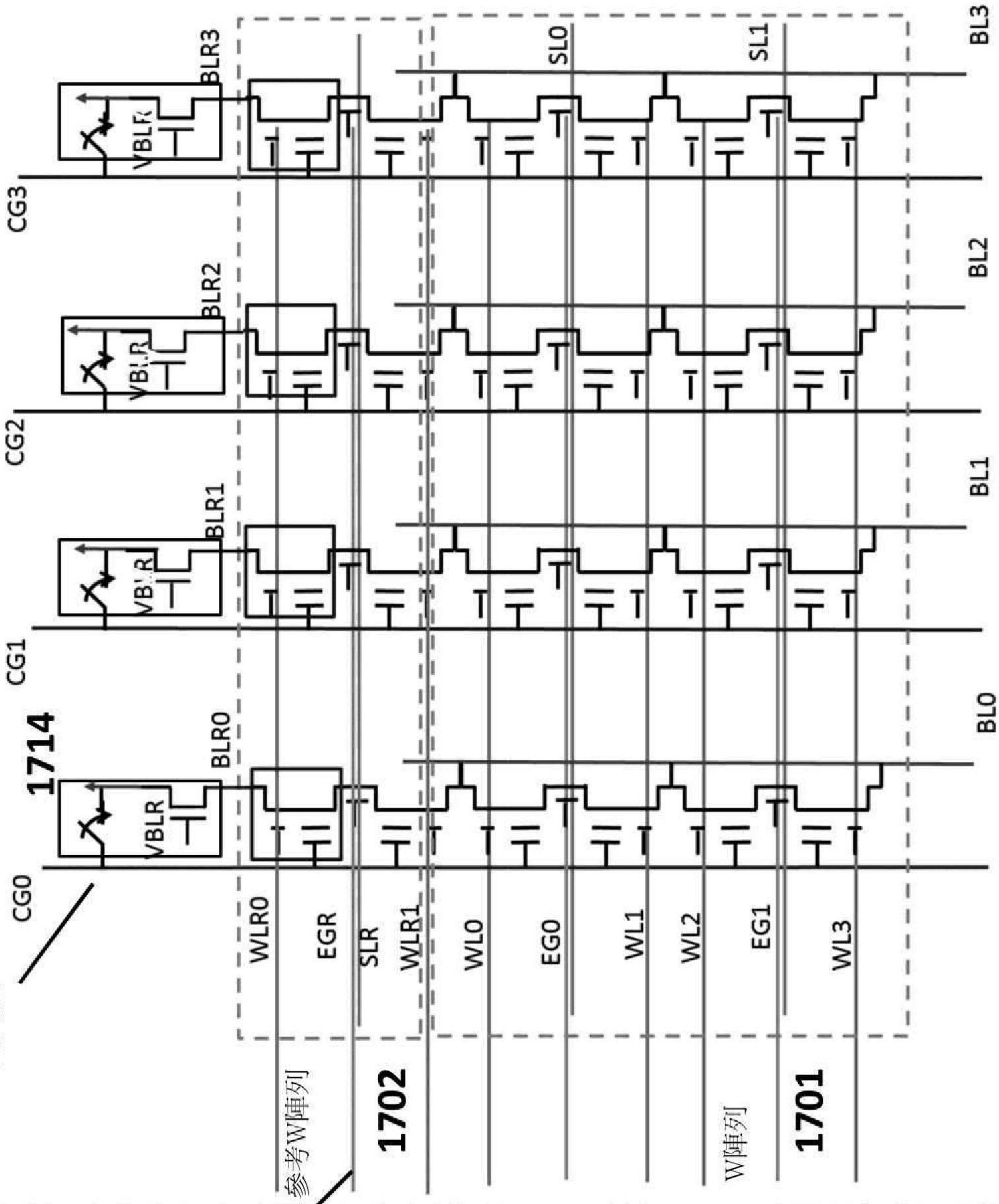
【圖15】

	WL	WL- 未經選擇	BL	BL- 未經選擇	CG	CG-未經選 擇相同磁區	CG- 未經選擇	EG	EG- 未經選擇	SL	SL- 未經選擇
讀取	1.0至2V	-0.5 V/ 0 V	0.6至2 V (Inuron)	0 V	0至2.6V	0至2.6V	0至2.6V	0至2.6V	0至2.6V	0 V	0 V
抹除	0 V	0 V	0 V	0 V	0 V	CGINH (4至9 V)	0至2.6V	5至12V	0至2.6V	0 V	0 V
程式化	0.7至1V	-0.5 V/ 0 V	0.1至1 uA	V _{inh} (1至2 V)	4至11V	0至2.6V	0至2.6V	4.5至5V	0至2.6V	4.5至5V	0-1 V

【圖16】

1700

1703



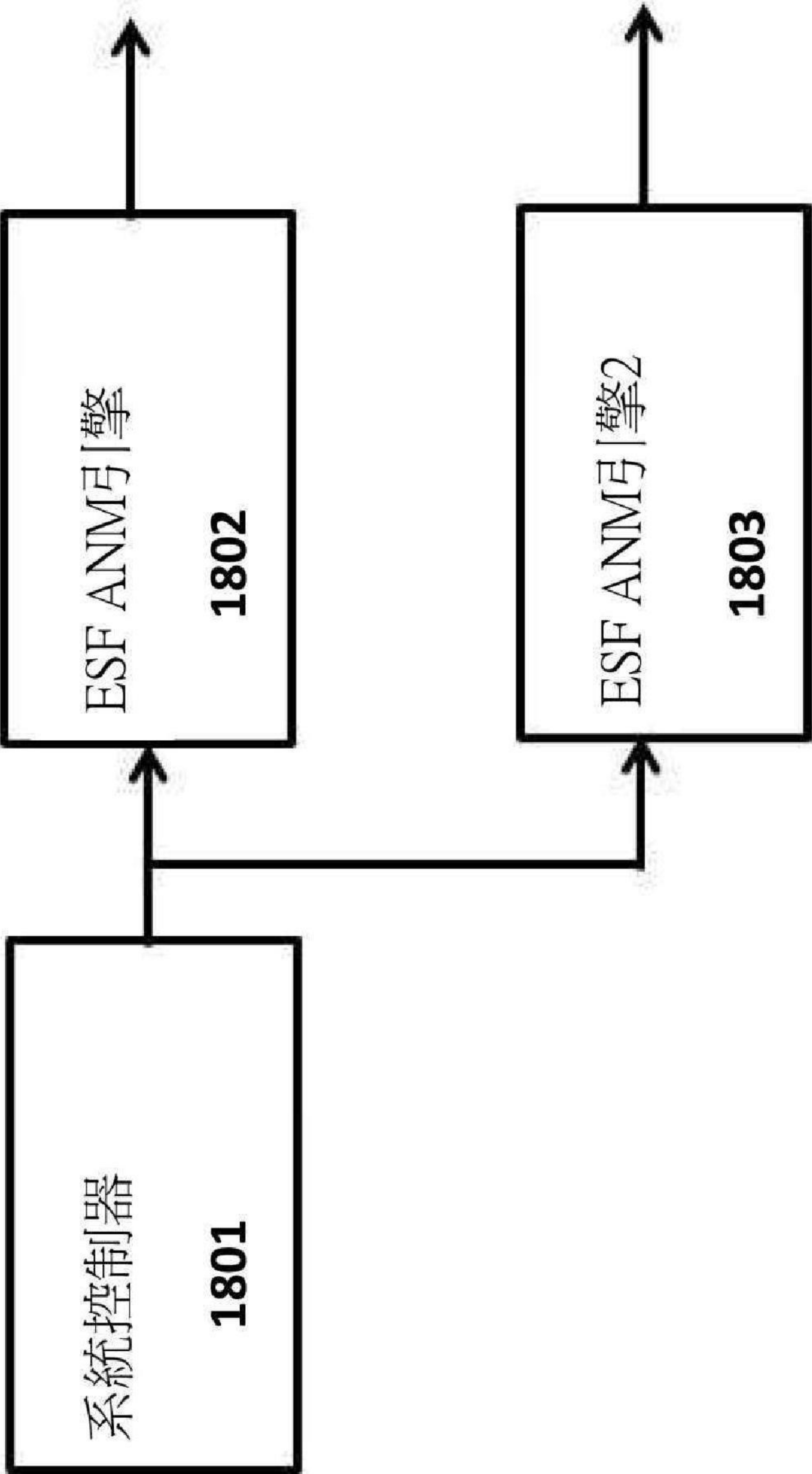
1704

1702

1701

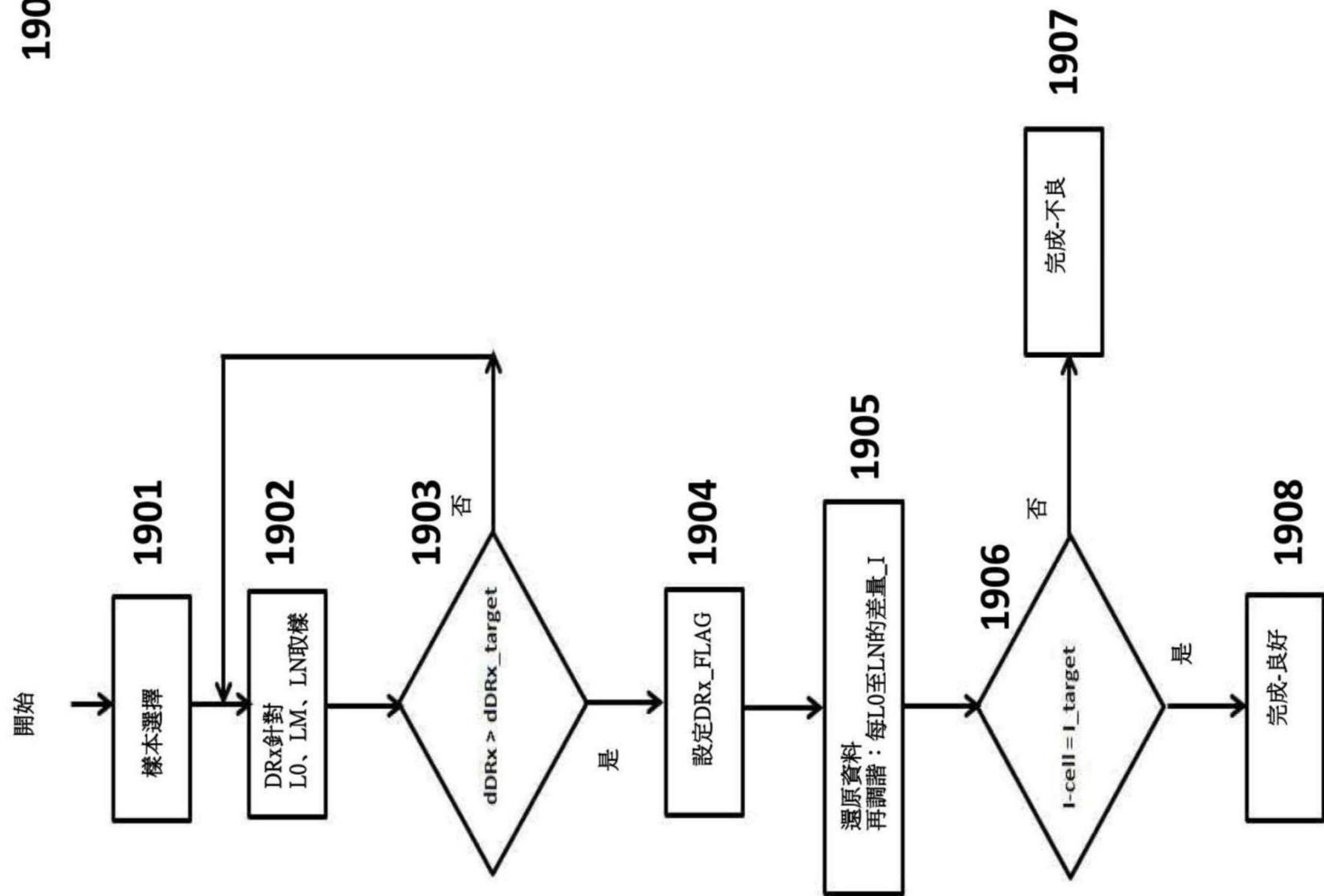
【圖17】

1800



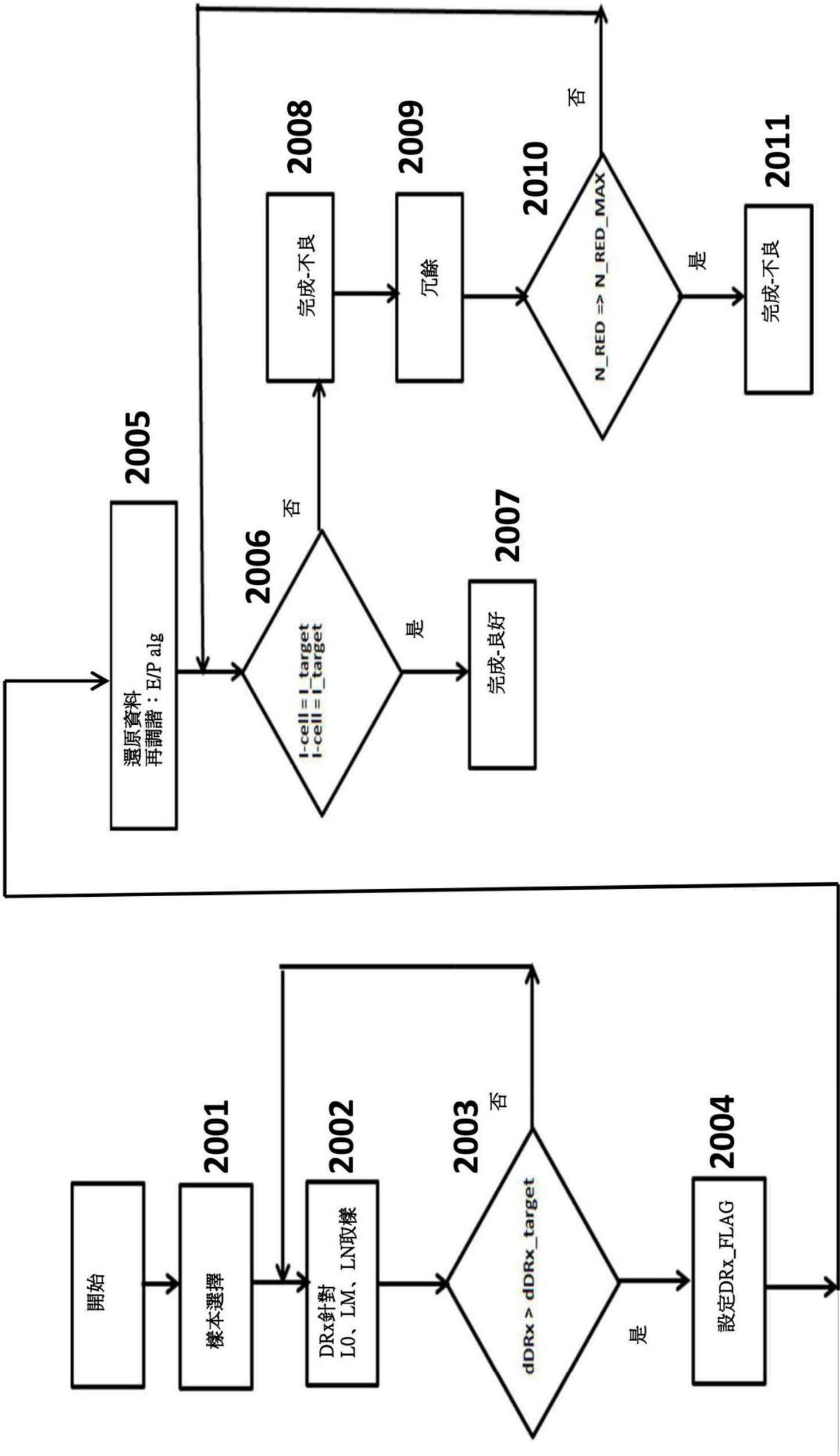
【圖18】

1900



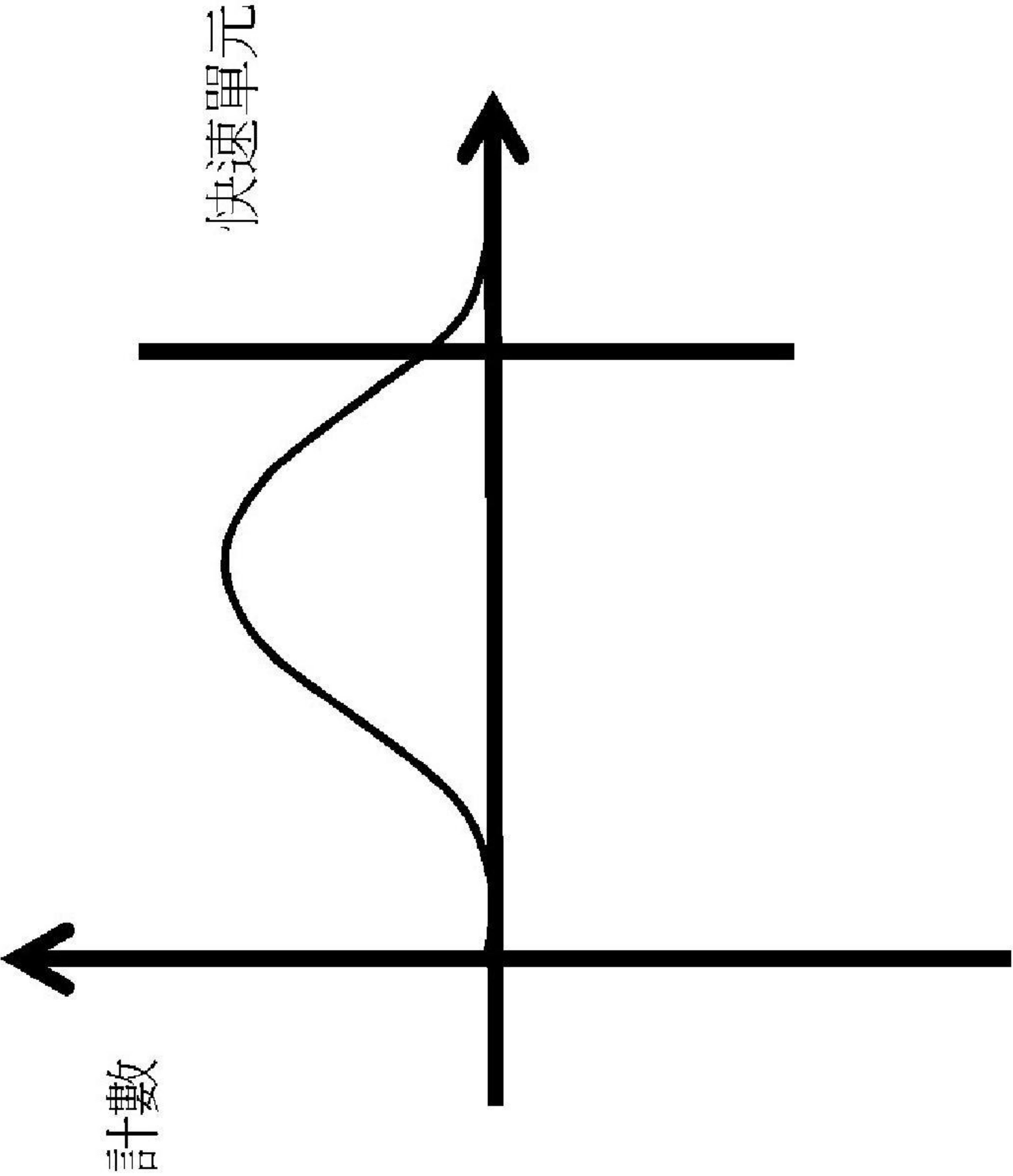
【圖19】

2000

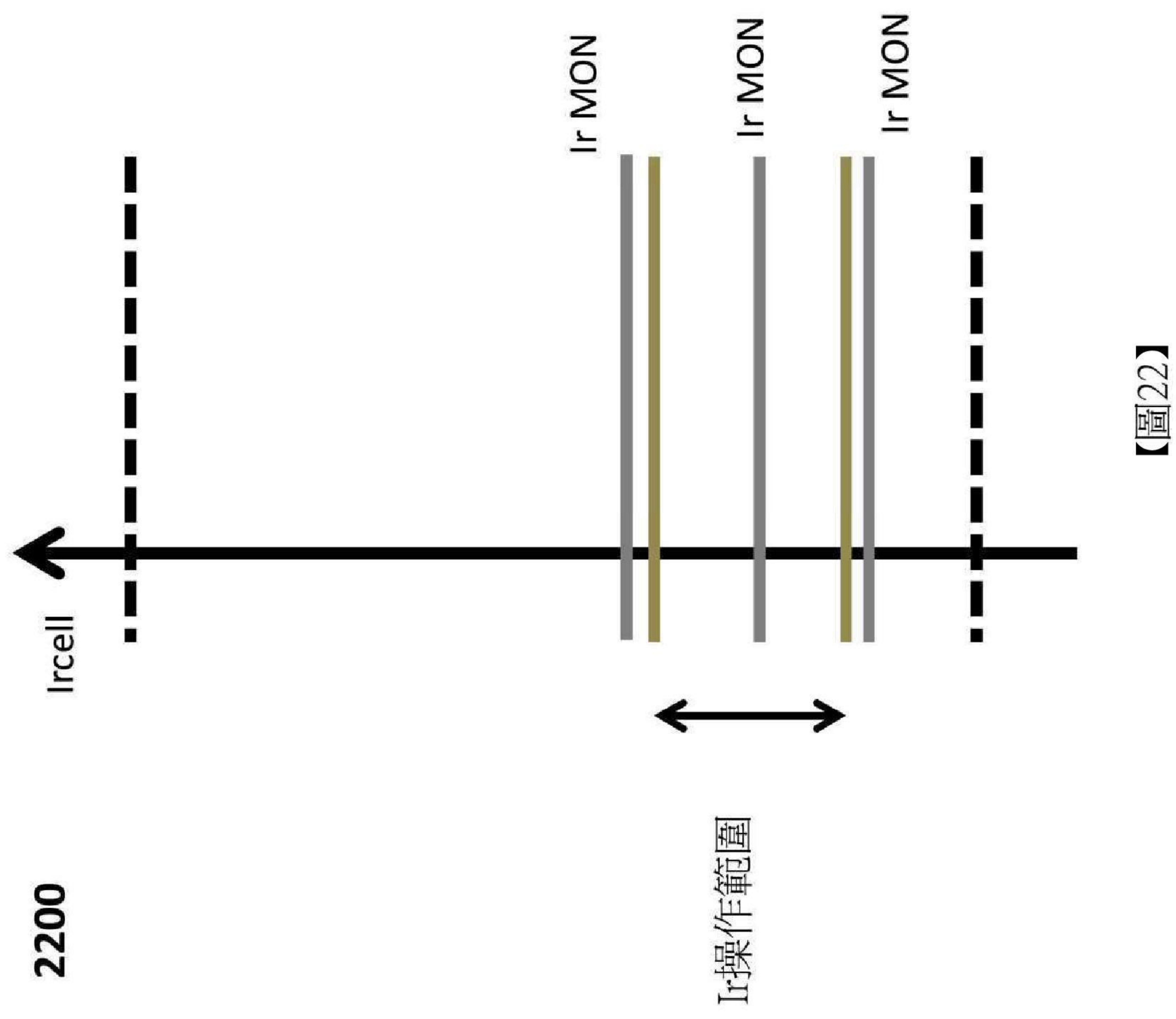


【圖20】

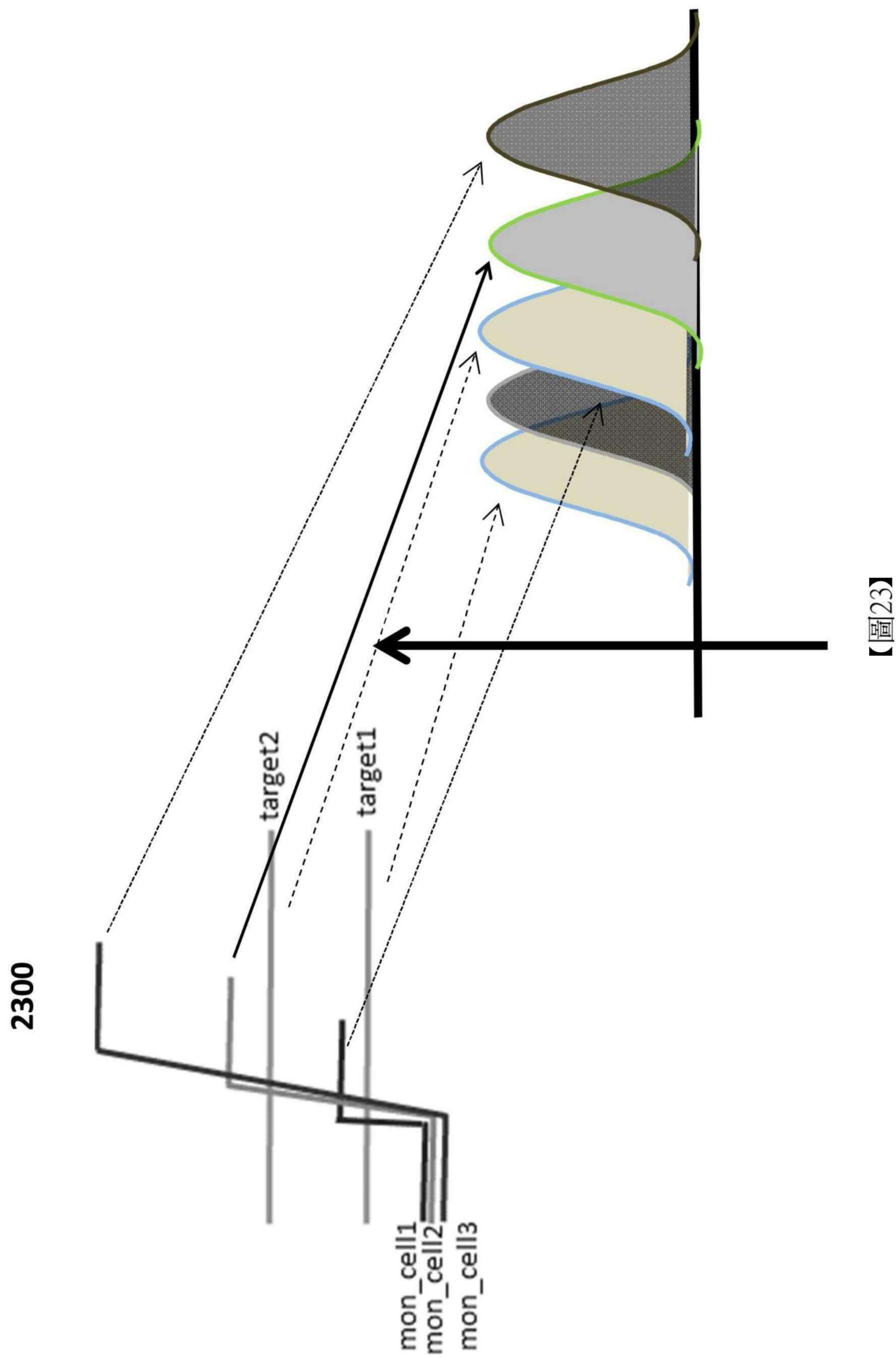
2100



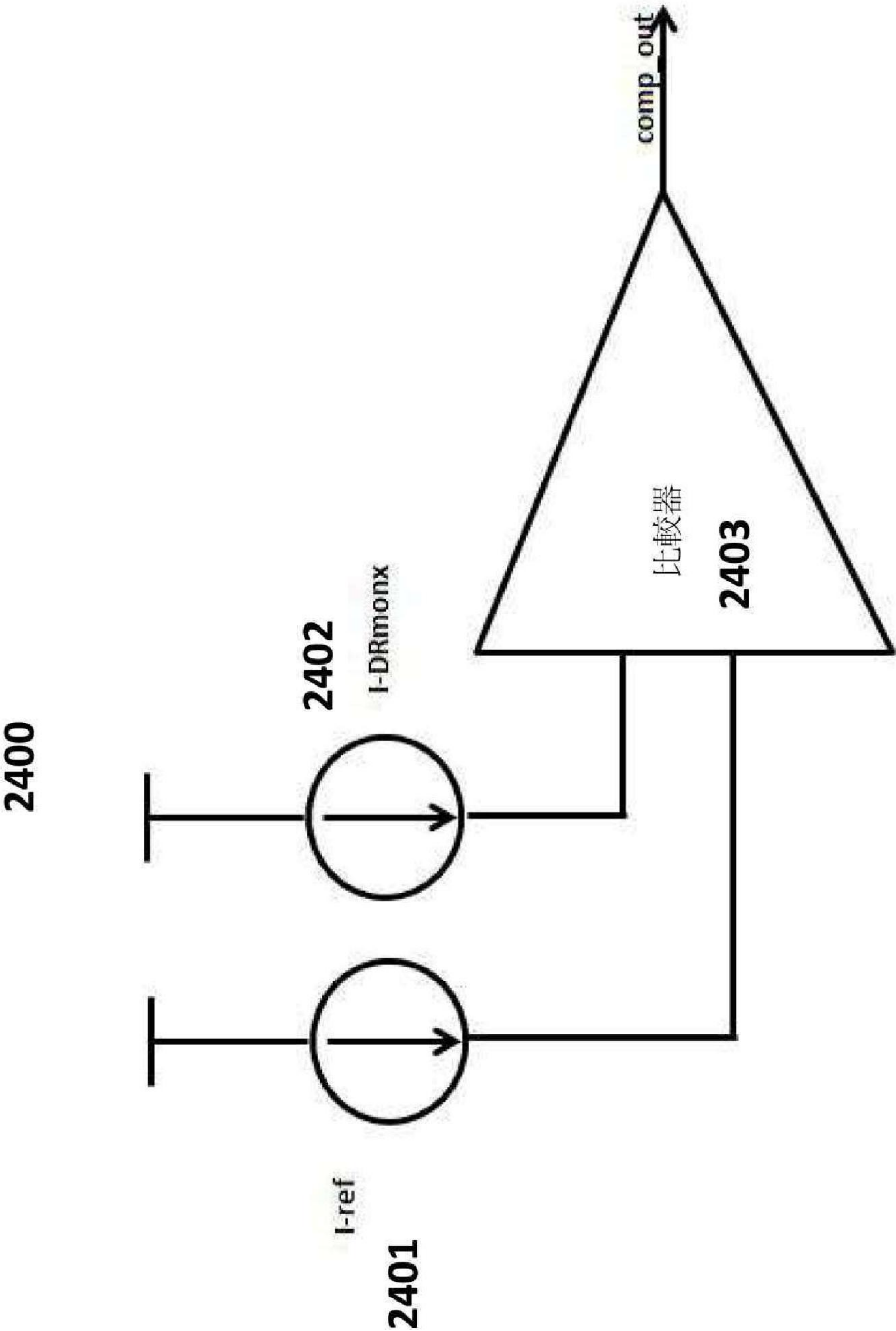
【圖21】



【圖22】

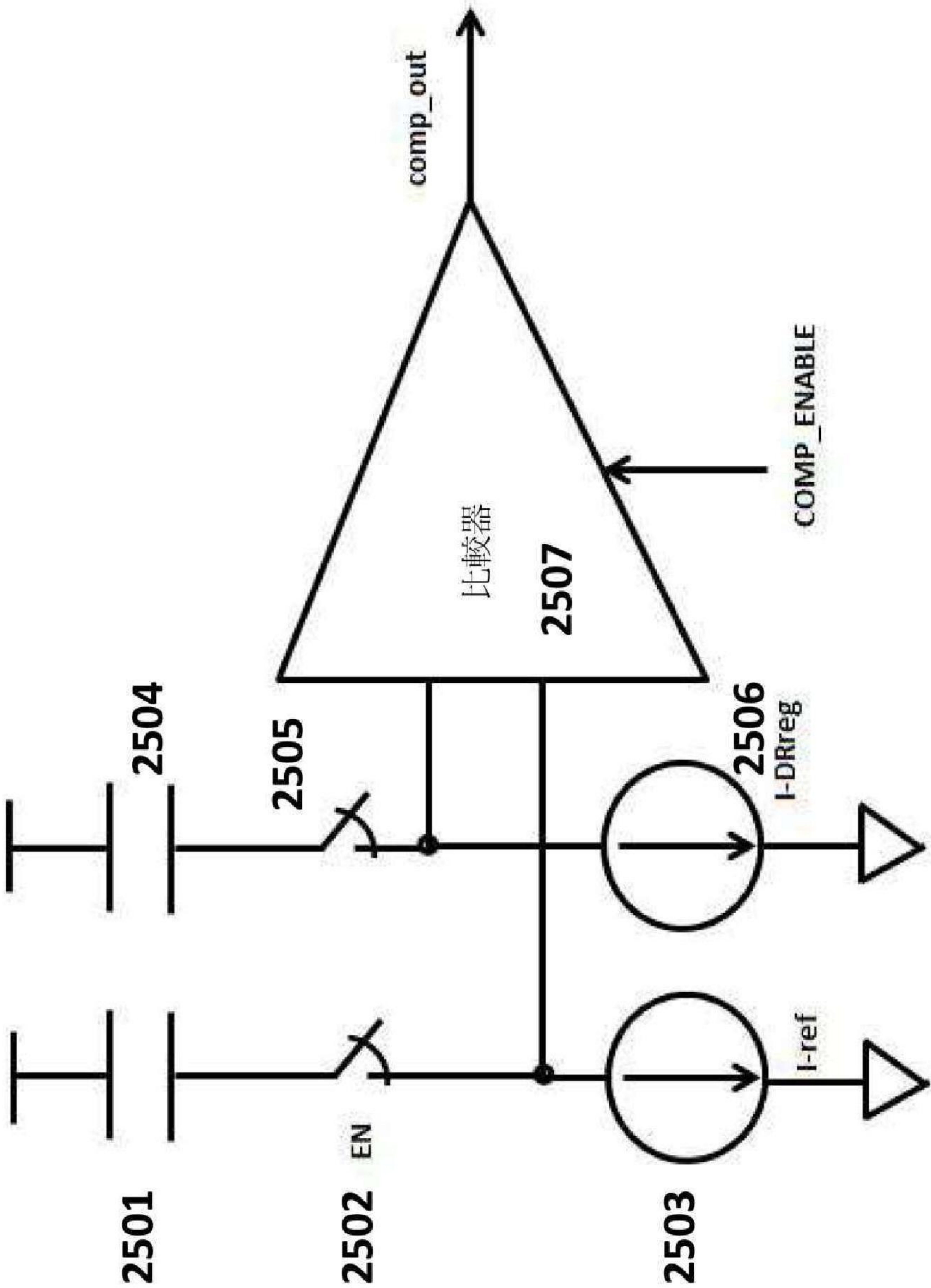


【圖23】



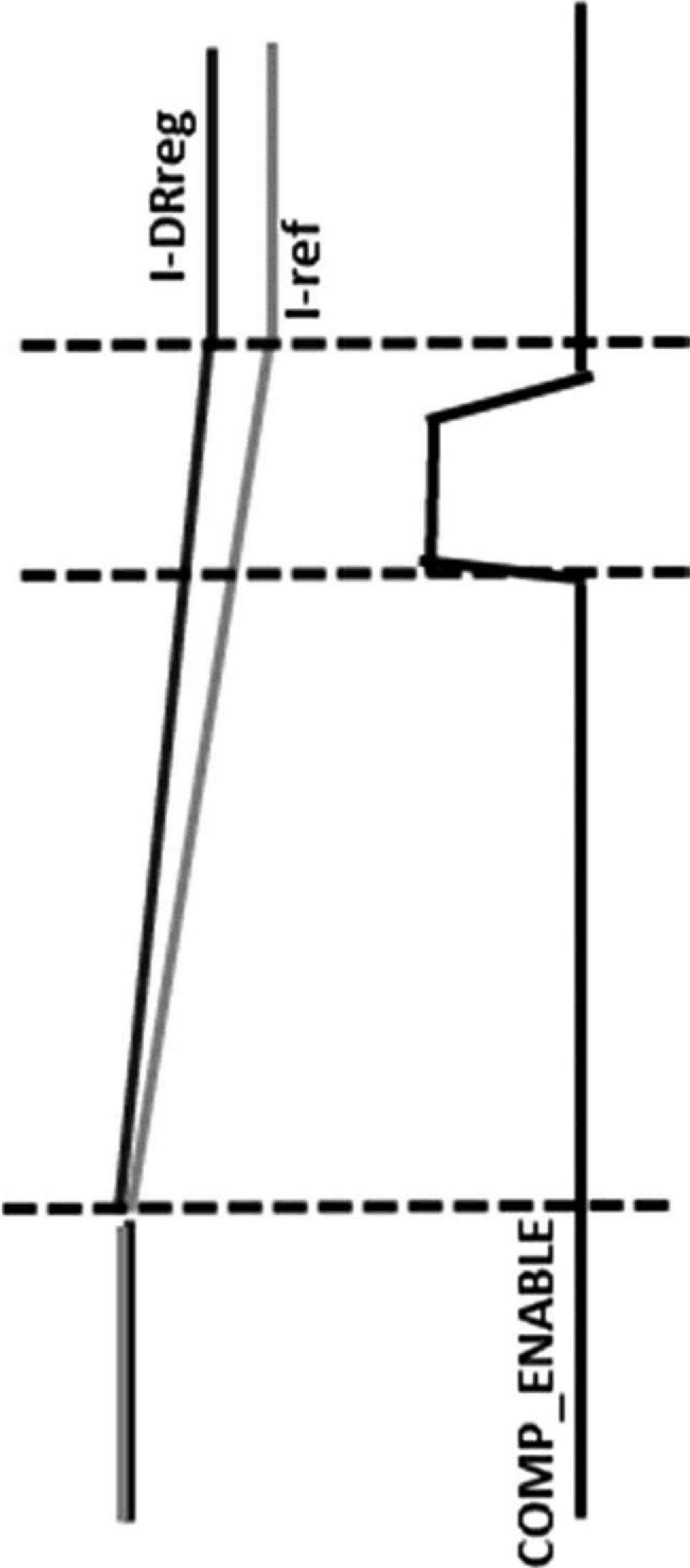
【圖24】

2500

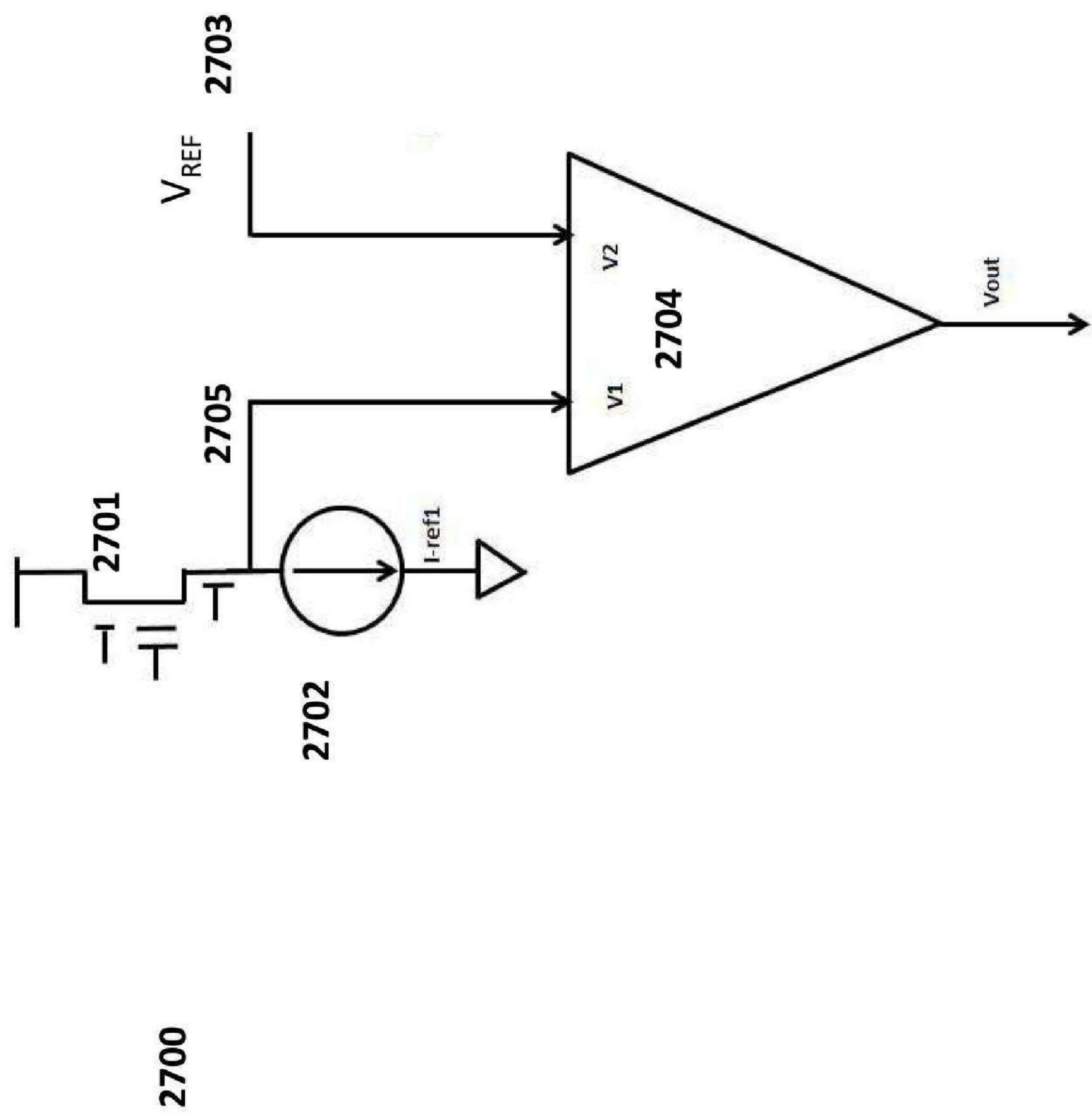


【圖25】

2600



【圖26】



【圖27】